

Pilot linkage of NHS Numbers for Babies data with birth registrations

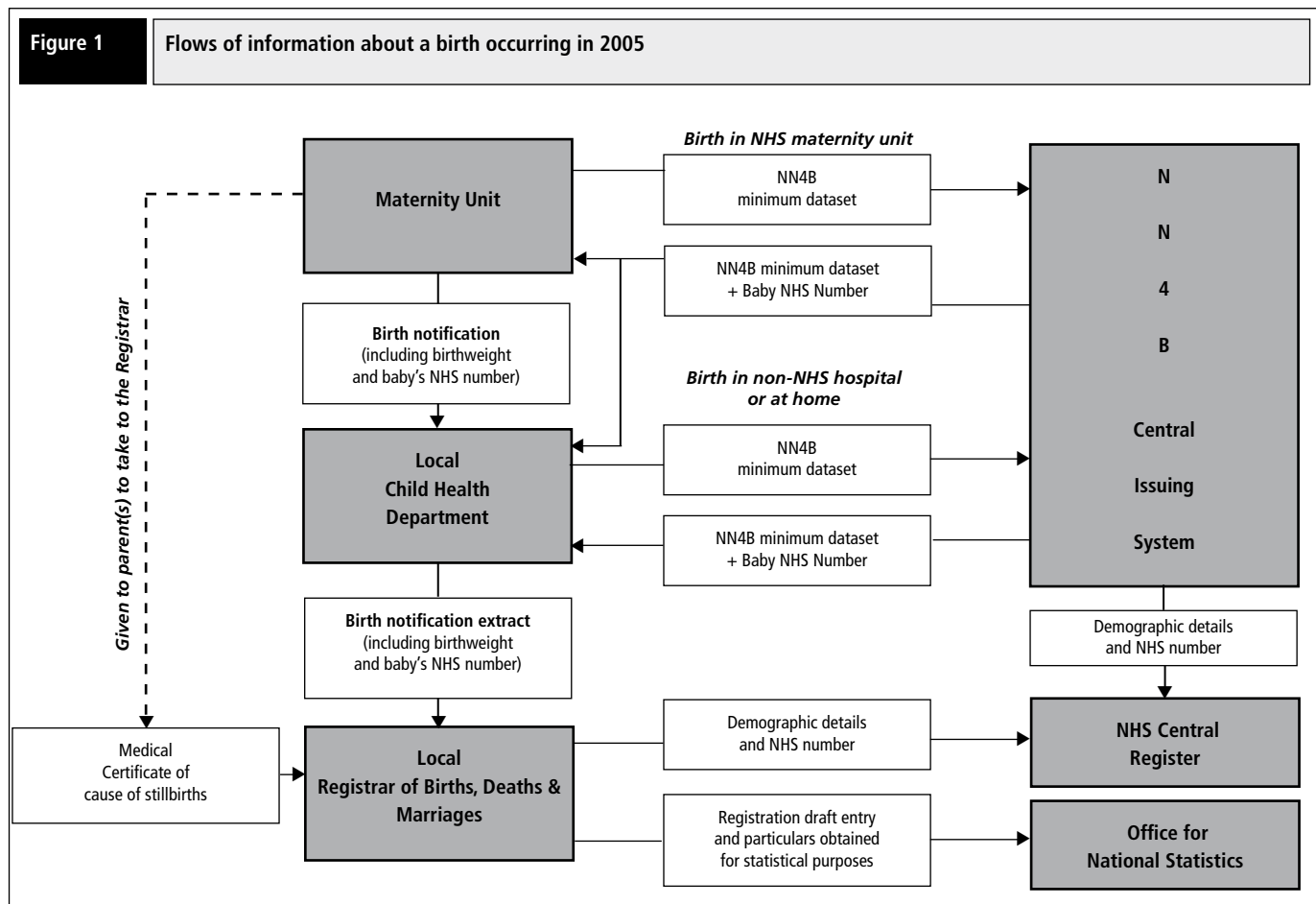
Lisa Hilder, Department of Midwifery, Institute of Health Sciences, City University; Kath Moser and Nirupa Dattani, Office for National Statistics; Alison Macfarlane, Department of Midwifery, Institute of Health Sciences, City University

Introduction

Information on gestational age at birth is of key importance as babies born preterm, before 37 completed weeks of gestational age, are at particular risk of morbidity and mortality in the early years of life.¹⁻³ Data about gestational age are needed to monitor trends in preterm birth, inform the development of programmes aimed at primary and secondary prevention of impairment and disability in childhood and to plan services for babies.

Information about births is required by various agencies and data flows in 2005 are summarised in Figure 1. Since 1974 gestational age at birth has been included in the birth notification that is sent by the birth attendant, usually the midwife, to the local child health department.⁴ To date ONS has not had access to this information about gestational age of live born babies as it is not recorded at registration in England and Wales. Although birth notification data have been successfully aggregated in regions where the same computer systems were used in all child health units,^{5,6} attempts to aggregate these data nationally have had less success. The opportunity to obtain gestational age information nationally was provided by the introduction of the NHS Numbers for Babies Service (NN4B) in 2002, as described below. This involves notifying births by sending a small set of data, including gestational age at birth, to the Central Issuing System that issues NHS numbers to babies born in England, Wales and the Isle of Man. As it was set up as an operational system to issue NHS numbers and was not intended to derive statistical information, the NN4B is a new and untested data source. A collaborative research project was set up in 2004 to link these data with birth and death registration data and develop gestational age statistics for England and Wales.

Information about gestational age is important but is not available for live births from registration data. It is, however, collected in NHS Numbers for Babies (NN4B) records. This project investigates the feasibility of linking NN4B data for births in the first quarter of 2005 with birth registration records. Overall 99.8 per cent of NN4B records linked with a registration record. Accuracy of linkage was questioned in 0.9 per cent. Live/stillbirth and multiple birth status were each differently classified in approximately 1 per 1,000 records. Discordance rates for other individual data items ranged from 0.3 per cent for date of birth to 12.9 per cent for postcode. Although needing further investigation, these results justify extending the linkage to the remainder of births in 2005. Linkage would be improved by retaining NHS numbers on stillbirth registration records and avoiding manual transfer of NHS numbers.



This article describes the initial stages of that work. The aims were to link NHS Numbers for Babies notification records and birth registration records and to use the linked records to assess the quality of the data from the two sources in terms of completeness and consistency. This was conducted as a pilot exercise using data for births occurring in the first three months of 2005.

Methods

Source data

Birth registration

By law, live births must be registered within six weeks of birth. The definition of a live birth, the legal basis, the process and a complete list of data items collected are described in detail elsewhere.^{7,8} The information is obtained, usually from one or both parents, by the local registrar. The local child health department (Figure 1) passes some information from the birth notification⁹ to the registrar to verify the birth. Since 1975 this has included the baby's birthweight and since October 2002 the NHS number.

The process for registration of stillbirths is similar to that for live birth except that registrars do not retain the NHS number for a stillbirth and the informant will also give the registrar a medical certificate of stillbirth completed by the attending midwife or doctor. This certifies the cause of death and includes an assessment of gestational age at birth and birthweight.

NHS Numbers for Babies Service (NN4B)

The NHS Numbers for Babies Service, which went live in October 2002, issues an NHS number to all registrable births as soon as possible after birth.^{10,11} This process replaced the issuing of NHS number at birth

registration. As the NHS number was mandated as the main identifier in communication between NHS agencies, and the development of a lifelong patient record, a mechanism was required for issuing an NHS number as soon as possible after birth. This change facilitated the better accounting of maternity resources, and the rapid collation of information and results, often from more than one centre, about newborn babies who are unwell.

Under this service a standardised set of information about each birth is notified electronically to the Central Issuing System which checks for duplication. If there is no existing record with the same details, the system returns a newly generated NHS number to the NHS agency notifying the birth. If a number has already been issued, the system will return a message indicating that a record with matching details already exists. The user is required to check this record against the details for the baby they have entered and use the existing number if a match is found. If no match is found, then the user will proceed to obtain the NHS number using the NN4B Child Health Browser, or remotely by telephone to the Number Issue Helpdesk if the browser is unavailable.

If the birth takes place in an NHS maternity unit, application for an NHS number is made by that unit (Figure 1). For births attended by independent midwives or in non-NHS hospitals application for an NHS number is usually made by the local child health department. When NHS midwives attend home births or births in non-maternity unit settings applications can be made either through the local maternity unit or the child health department.

Records on the Central Issuing System are held for six months and then deleted. When the NHS number is issued, a copy of the information is sent to the child health system local to the place of birth. Daily messaging sends a limited dataset, including only those items sent by the NHS for birth registration, to the NHS Central Register.

Data used in this analysis

An NN4B record for each NHS number issued to a newborn baby from 1st January 2005 was sent to ONS. For this pilot, records relating to babies born during the three month period from 1st January to 31st March 2005 inclusive were selected from the NN4B and registration data. The registration data are from a provisional extract taken in November 2005. Selected data items available for statistical analysis from each source are listed in Table 1. Approval for the use of these data sources for linkage and the production of statistical data was given by the North East London Ethics Committee and the Patient Information Advisory Group.

Table 1 Selected data items available from NHS Numbers for Babies records and birth registration records

Data items	Data sources	
	NHS Numbers for Babies	Birth registration
Baby NHS number	+	+
Mother's NHS number	+	+
Birth date of baby	+	+
Delivery time	+	+
Birthweight	+	+
Gestational age at stillbirth	+	+
Gestational age at live birth	+	+
Sex of baby	+	+
Number of babies born	+	+
Suspected congenital anomaly – Yes / No	+	+
Live or stillbirth	+	+
Cause of death (stillbirths)		+
Previous children to mother within marriage ¹		+
Mother's date of birth	+	+
Baby's postcode of usual residence	+	+
Social class from occupation of mother		+
Marital status of mother		+
Joint or sole registration of birth		+
Social class from occupation of father ²		+
Ethnic category of baby	+	
Country of birth of mother		+
Country of birth of father ²		+
Place of birth – (name of hosp, unit)	+	+
Delivery place type (e.g. NHS hosp, home)	+	+
GP practice code	+	

1. Where the child is born within marriage.

2. Available if parents are married or jointly register the birth.

Record linkage

The NHS number, being a unique direct identifier and included in both datasets, was the primary means to link records. By definition, the NN4B data always have an NHS number. If registration records had no NHS number, indirect identifiers had to be used for matching. In order to maximise electronic linkage and minimise time-consuming manual linkage, the variables chosen for linkage were those with the largest number of different values. An iterative process was used, applying baby's date of birth and different combinations of birthweight, mother's date of birth, and postcode of usual residence. Only records remaining unmatched at the end of each matching cycle were entered into the next stage of the matching process.

As the linkage progressed, the number of records remaining unmatched became smaller, thus reducing the chance of there being two babies with the same characteristics. For this reason, matching criteria were made less stringent, requiring two rather than three data items to match. In addition the mother's year of birth was used in place of her date of birth and the first part of the postcode was used rather than full postcode.

The final linkage algorithm involved 15 stages. The combinations of variables and the order in which they were used is shown in Figure 2.

Stillbirths and live births were linked separately to allow the progress of linkage in these two groups to be viewed more easily. Live or stillbirth status acts as a matching variable in stages 2–12. Only singleton births were entered into stages 6–12, where birthweight was omitted, in order to avoid false matches. All live and stillbirth records remaining unmatched were entered into stage 13. The process culminated in the manual examination of the remaining unmatched records, sorted by date of birth. At each stage the number of records entered was checked against the total number of records which were matched and unmatched. A variable denoting the stage of matching was added to the data.

Data quality

As the established source of birth data, birth registration is considered to include all babies because of the civil uses of the birth and stillbirth certificates issued. The data collected at registration are subject to a series of quality control measures.⁷

Assessment of the quality of NN4B data was based on the coverage, completeness and consistency of NN4B data relative to birth registration data. Coverage refers to the extent to which data are available for all registrable births. A crude method compares the number of NN4B records with those for registrations. One-to-one record linkage is a powerful test that the two sources contain records for the same individuals. Completeness refers to the extent to which individual data items have non-missing data. Null and dummy values were included with missing values.

In linked data, the values for data items common to both sources can be compared. If the values from sources for a given variable are identical, the data are described as concordant. If they differ, the data are discordant. Measurement of the extent of data consistency can be used to comment on data quality in one source relative to another if the data item was not used for linkage, if there is non-missing data in data items from each source and if there is no doubt about the accuracy of the record linkage.

Doubt is raised about the accuracy of record linkage if identical values are not seen in the same data items from both sources. The exception is postcode which can change if the mother moves house before she registers the birth. As some recording errors are to be expected in routine data, allowance is made for data discordance in one data item, but there is serious doubt about the linkage if three or more items of data are discordant.

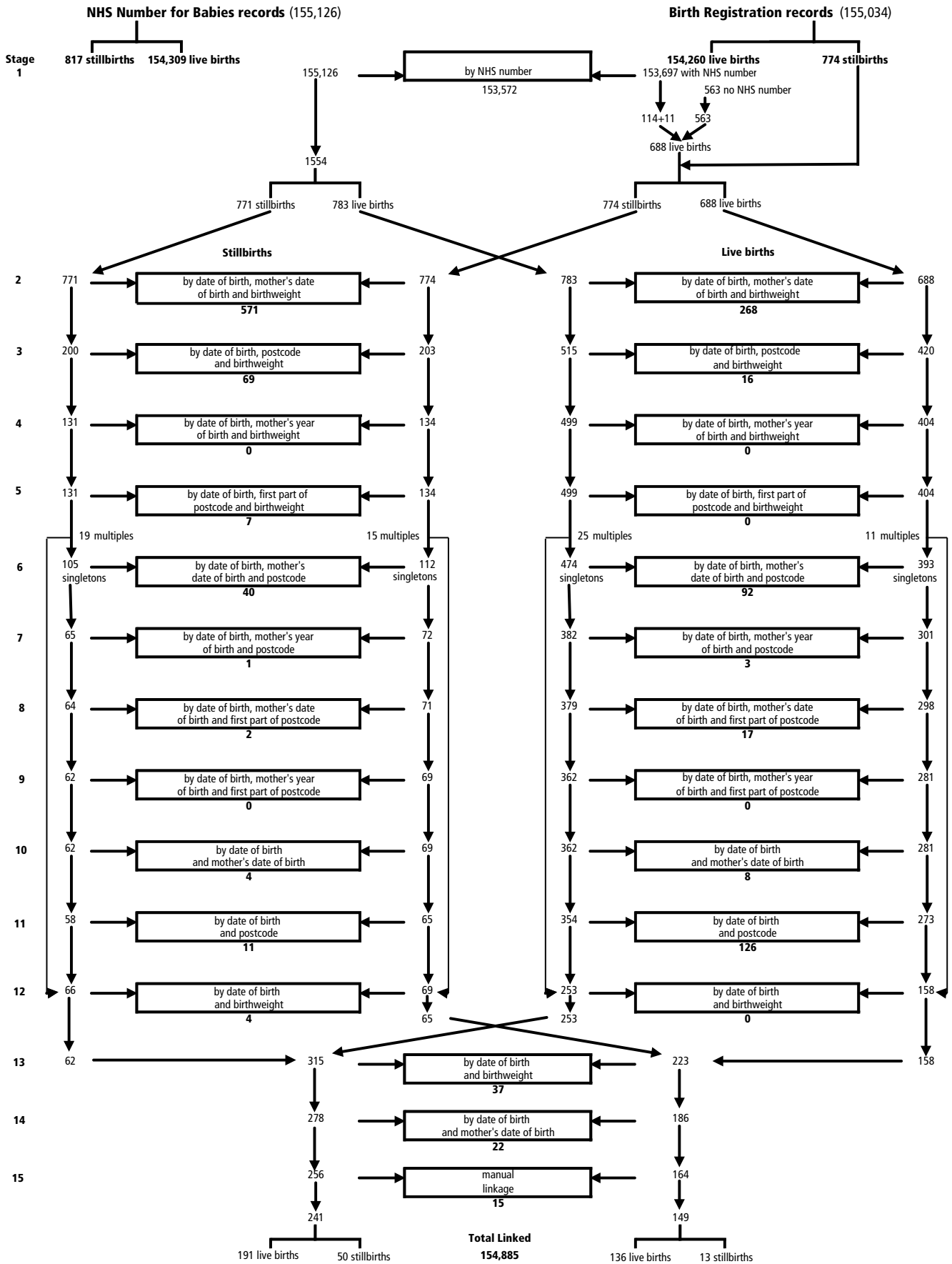
Data processing, data linkage and analysis were carried out using SAS software.¹²

Table 2 Total number, and number linked, of NHS Numbers for Babies records and birth registration records, by live/stillbirth status

		NHS Numbers for Babies	Birth registration
Total records			
All births		155,126	155,034
Stillbirths		817	774
Live births		154,309	154,260
Number and % linked			
All births	Number	154,885	154,885
	%	99.8	99.9
Stillbirths	Number	767	761
	%	93.9	98.3
Live births	Number	154,118	154,124
	%	99.9	99.9

Figure 2

The record linkage process showing the variables used for matching, number of records entered, matched and remaining at each stage of the linkage



Results

Record linkage

For babies born in England and Wales from January 1st to March 31st 2005 there were 155,126 NHS Numbers for Babies records and 155,034 birth registration records after one duplicate record was identified and removed from each dataset (Table 2). Of these, there were 817 NN4B records and 774 registrations relating to a stillbirth. Using the algorithm set out in Figure 2 NN4B and registration records were linked as described in detail below.

Stage 1 Matching by NHS number

A total of 153,572 records (99.1 per cent of registration records and 99.0 per cent of NN4B records) were matched using NHS numbers. There were 114 live birth registrations that failed to link in this stage despite having an NHS number. The possibility that registration records might have the same NHS number was not initially considered, but this stage of linkage identified 11 pairs of records for different babies with the same NHS number. In each case, the 'true' registration record was identified as the one with identical data in the common fields on the NN4B record with that NHS number. The NHS numbers for the other registration records were recoded to missing. These 11, plus the 563 live birth registrations with no NHS number, the 114 that failed to link on NHS numbers, and the 774 stillbirth registrations, were entered into the second stage. The 1,554 NN4B records that did not link on NHS numbers were also entered into the second stage.

Stages 2–15 Matching by other variables

Stillbirths and live births were considered separately throughout stages 2–12 inclusive. The second stage of linkage attempted to match records using the baby's date of birth, mother's date of birth and baby's birthweight. Among the stillbirths, of the 771 NN4B and 774 registration records entered at this stage, 571 records linked using the three variables; they represent 74.1 per cent of the NN4B records entered and 73.8 per cent of the registrations. Among the live births, only 268 records, representing 34.2 per cent of the NN4B records and 39.0 per cent of the registrations entered, were linked at this stage. In the third stage, which used date of birth, postcode and birthweight for matching, 69 stillbirths and 16 live births linked. In stages 4–12 a further 69 stillbirths and 246 live births were linked. Higher numbers of matches occurred after birthweight was removed from the matching criteria at stage 6 and mother's date of birth was removed at stage 11. In part, this probably reflects the extent to which data are missing from these fields (Table 4). In stages 13 and 14, where live births and stillbirths were no longer treated separately, a further 59 records were matched using date of birth and either birthweight or mother's date of birth. In the final stage (15), a manual search for matching records linked 15 records from each data source. At the end of the matching process, 241 NN4B records and 149 registration records remained unlinked.

Overall 154,885 records, representing 99.8 per cent of NN4B records and 99.9 per cent of registrations, were linked (Table 2). It was possible

Table 3

Number of linked records with discordant data¹ by number of discordant data items, stage at which linkage occurred, discordance in live or stillbirth status and discordance in multiple birth status

	Linkage using:				Discordance in live/stillbirth status				Discordance in multiple birth status				All records	
	NHS Number		other variables		No		Yes		No		Yes			
	Number	%	Number	%	Number	%	Number	%	Number	%	Number	%	Number	%
No data item with discordant data¹	128,189	83.5	984	74.9	129,109	83.4	64	74.4	129,039	83.4	134	79.3	129,173	83.4
One data item with discordant data	24,073	15.7	306	23.3	24,357	15.7	22	25.6	24,347	15.7	32	18.9	24,379	15.7
NHS number	0		25		25		0		25		0		25	
Date of birth	316		3		319		0		319		0		319	
Birthweight	1,795		65		1,856		4		1,853		7		1,860	
Mother's birth date	3,237		38		3,274		1		3,273		2		3,275	
Postcode	18,064		167		18,215		16		18,208		23		18,231	
Sex	661		8		668		1		669		0		669	
Two data items with discordant data	1,264	0.8	21	1.6	1,285	0.8	0	0.0	1,284	0.8	1	0.6	1,285	0.8
Birthweight/NHS number	0		3		3		0		3		0		3	
Birthweight/date of birth	12		0		12		0		12		0		12	
Mother's birth date/NHS number	0		2		2		0		2		0		2	
Mother's birth date/date of birth	13		0		13		0		13		0		13	
Mother's birth date/birthweight	75		1		76		0		76		0		76	
Postcode/NHS number	0		3		3		0		3		0		3	
Postcode/date of birth	57		0		57		0		57		0		57	
Postcode/birthweight	345		8		353		0		352		1		353	
Postcode/mother's birth date	616		2		618		0		618		0		618	
Sex/date of birth	2		0		2		0		2		0		2	
Sex/birthweight	14		0		14		0		14		0		14	
Sex/mother's birth date	15		1		16		0		16		0		16	
Sex/postcode	115		1		116		0		116		0		116	
Three data items with discordant data	38	0.0	2	0.2	40	0.0	0	0.0	39	0.0	1	0.6	40	0.0
Four data items with discordant data	7	0.0	0	0.0	7	0.0	0	0.0	6	0.0	1	0.6	7	0.0
Five data items with discordant data	1	0.0	0	0.0	1	0.0	0	0.0	1	0.0	0	0.0	1	0.0
<i>Two or more items with discordant data</i>	<i>1,310</i>	<i>0.9</i>	<i>23</i>	<i>1.8</i>	<i>1,333</i>	<i>0.9</i>	<i>0</i>	<i>0.0</i>	<i>1,330</i>	<i>0.9</i>	<i>3</i>	<i>1.8</i>	<i>1,333</i>	<i>0.9</i>
<i>Two or more items with discordant data, excl postcode</i>	<i>177</i>	<i>0.1</i>	<i>9</i>	<i>0.7</i>	<i>186</i>	<i>0.1</i>	<i>0</i>	<i>0.0</i>	<i>184</i>	<i>0.1</i>	<i>2</i>	<i>1.2</i>	<i>186</i>	<i>0.1</i>
All records	153,572	100.0	1,313	100.0	154,799	100.0	86	100.0	154,716	100.0	169	100.0	154,885	100.0

1. Data items assessed for discordance included NHS number, date of birth, birthweight, mother's date of birth, postcode, sex.

to link 153,572 of them using NHS numbers, while the remaining 1,313 were linked using other variables. A smaller proportion of stillbirths than live births linked (98.3 per cent of stillbirth registrations and 93.9 per cent of stillbirth NN4B records) (Table 2).

Accuracy of data linkage

The number of data items with discordant data within the same linked record are tabulated in Table 3 according to the stage at which the records were linked. There was no discordance in any of the data items common to each source in 83.4 per cent of all linked records. This was similar to the 83.5 per cent complete concordance rate for records linked using NHS numbers, but higher than the 74.9 per cent of records with complete concordance where linkage was achieved using indirect identifiers. There is some doubt about the accuracy of record linkage for the 0.9 per cent of all linked records that have discordant data in two or more fields. If discordant postcodes are ignored the proportion of records with doubtful linkage is reduced to 0.1 per cent. In 0.03 per cent of all linked records there is serious doubt about the accuracy of record linkage. In linkage using indirect identifiers there were 1.8 per cent of records where there is doubt about the accuracy of record linkage, or 0.7 per cent of records if discordance in postcode is not counted, and serious doubt for 0.15 per cent.

Data quality

Missing data

Table 4 shows the number and percentage of the linked records which had missing or discordant data in each of the common data fields. A valid date of birth and sex of baby are included in all records. In 0.8 per cent of linked records, the mother’s date of birth was missing from the registration data while this was missing from only 0.1 per cent of NN4B records. A higher proportion of records had NN4B birthweight data missing (0.4 per cent) than registration birthweight data (0.2 per cent). One in forty linked records were missing postcode from the NN4B data, while less than 0.01 per cent of records were missing postcode from registration data. Very few records had data missing from both NN4B and registration sources for the same item.

Table 4 Number of linked records with missing or discordant data items in common data fields

Data items	Missing data by data source				Discordant data
	NN4B records	Registration records	Either source	Both sources	
NHS number	0	1,274	1,274	0	35
Date of birth	0	0	0	0	414
Sex of baby	0	0	0	0	836
Birthweight	658	382	980	30	2,354
Mother’s date of birth	78	1,306	1,380	2	4,042
Postcode	3,808	41	3,841	4	19,421
	% of all linked records (154,885)				
NHS number	.	0.8	0.8	.	0.0
Date of birth	0.3
Sex of baby	0.5
Birthweight	0.4	0.2	0.6	0.02	1.5
Mother’s date of birth	0.1	0.8	0.9	0.00	2.6
Postcode	2.5	0.0	2.5	0.00	12.5

Discordant data in individual common data items

For a given data item, discordance was assessed only where information was recorded on both registration and NN4B information records. In 83.4 per cent of linked records there were no discordant data for any data item (Table 3). As shown in Table 4, date of birth was discordant in 0.3 per cent of records, sex in 0.5 per cent, birthweight in 1.5 per cent, mother’s date of birth in 2.6 per cent and postcode in 12.5 per cent of

records. More than one pair of data items may be discordant in any one record. For baby’s sex and mother’s age in five year bands, there was no discernable difference in the percentage distributions of data from the two sources. Differences in the distributions of birthweight data are described below. NHS numbers were discordant in 35 records.

Discordance in multiple birth status and live or stillbirth status

Looking at multiple births, 4,462 were identified as multiple births in both sources. Of the 4,534 NN4B records identified as a multiple birth, 72 linked with a singleton registration record, while of the 4,559 multiple birth registration records, 97 linked with a singleton NN4B record (Table 5). This results in discordance in data about multiple birth status in 169 (0.11 per cent) linked records. Only 3 of these 169 records had discordant data in more than one other replicated data field (Table 3).

Table 5 Discordance in multiple birth status in linked records

NN4B information	Registration information		
	Singleton	Multiple	Total
Singleton	150,254	97	150,351
Multiple	72	4,462	4,534
Total	150,326	4,559	154,885

There was discordance in live or stillbirth status in 86 records (Tables 3 and 6), representing 0.06 per cent of linked records. Among the 767 NN4B records identified as a stillbirth, 46 linked with a live birth registration while 27 of the 761 stillbirth registrations linked with a live birth notification and 13 with a NN4B record classified as ‘dead’. The option ‘dead’ on NN4B records allows the clinician to indicate that a baby born alive had died before the notification was made (no such category is available in the registration records). These births are counted as live born in the analysis presented in this paper. None of the 86 records with discordant live or stillbirth status had discordant data in more than one other replicated data field (Table 3).

Table 6 Discordance in live or stillbirth status in linked records

NN4B information	Registration information		
	Live birth	Stillbirth	Total
Live	154,032	27	154,059
Dead	46	13	59
Still birth – antepartum	31	500	531
Still birth – indeterminate	4	147	151
Still birth – intrapartum	11	74	85
Total	154,124	761	154,885

Discordance in birthweight

Birthweights in 500g groups from NN4B data were compared with grouped birthweights from birth registration data (Table 7). Overall 98.6 per cent of linked records were in the same 500g birthweight group according to both data sources. However this generally good agreement masks some important differences. Many more records with a stated birthweight (453) are in the under 500g group according to the NN4B data than the birth registration data (179). All the other 500g groups below 5,000g contain more records according to the registration data than the NN4B data. The differences are proportionally larger in the groups under 2,500g. The overall discordance is greater when exact (ungrouped) birthweights are compared (1.5 per cent of linked records (Table 4)) than when birthweights in 500g groups are compared as here (1.4 per cent of linked records).

Table 7 Birthweight distributions of linked records using birthweight data from registration and NHS Number for Babies records

Birthweight (g)	NHS Numbers for Babies data													Total	%
	<500	500-999	1,000-1,499	1,500-1,999	2,000-2,499	2,500-2,999	3,000-3,499	3,500-3,999	4,000-4,499	4,500-4,999	5,000-5,499	5,500 and over	Not stated		
<500	162	3	.	2	.	.	2	.	3	1	.	.	6	179	0.1
500-999	7	823	5	1	.	1	.	4	.	.	.	2	36	879	0.6
1,000-1,499	5	2	1,218	3	1	1	5	1	52	1,288	0.8
1,500-1,999	14	1	6	2,366	5	2	2	1	1	.	.	.	51	2,449	1.6
2,000-2,499	17	1	4	5	7,256	62	86	44	13	3	1	.	42	7,534	4.9
2,500-2,999	55	1	1	2	25	26,237	31	62	2	1	1	.	97	26,515	17.1
3,000-3,499	89	3	2	2	36	40	54,717	113	22	1	.	7	171	55,203	35.6
3,500-3,999	68	4	1	.	6	50	119	43,173	9	6	.	3	141	43,580	28.1
4,000-4,499	18	2	4	3	1	3	24	8	14,223	5	1	3	29	14,324	9.2
4,500-4,999	9	.	.	.	1	.	4	9	11	2,271	1	.	3	2,309	1.5
5,000-5,499	1	1	.	.	1	.	1	2	.	.	216	.	.	222	0.1
5,500 and over	.	1	1	.	.	1	18	.	21	0.0
Not stated	8	9	2	8	21	57	119	97	27	3	1	.	30	382	0.2
Total	453	851	1,243	2,392	7,353	26,453	55,110	43,515	14,311	2,291	222	33	658	154,885	
%	0.3	0.5	0.8	1.5	4.7	17.1	35.6	28.1	9.2	1.5	0.1	0.0	0.4		

Discordance in gestational age for stillbirths

Information about gestational age at birth is available from both registration and NN4B records for 749 of the 761 linked stillbirth registrations. Table 8 shows that, with few exceptions, there is good agreement between the values for gestational age recorded in the two sources. Gestational ages agreed in 81.8 per cent of these records and differed by one week in 13.5 per cent; only in 4.7 per cent was there a difference of two weeks or more.

Discussion

The linkage of NHS Numbers for Babies data with birth registration records described here was conducted as a pilot study into the feasibility of the process, using data on births in the first quarter of 2005. Importantly, because of the timing of this work, a provisional extract of registration records had to be used. This early extract would not have been through the full range of data cleaning and quality checks that are

performed before the data are finalised. Despite using this provisional extract we successfully linked most birth records, and found data items common to both sources to be in agreement in most cases. More specifically, using the NHS number, it was possible to link 99.0 per cent of NHS Numbers for Babies records for births in England and Wales in the first quarter of 2005 with a birth registration record. A further 0.8 per cent of NN4B records could be linked using other variables including date of birth and combinations of birthweight, mother’s date of birth and postcode of usual residence. In most instances the linkages were shown to be accurate but in 0.9 per cent of linked records there were discordancies between the two sources in two or more of NHS number, date of birth, birthweight, mother’s date of birth, postcode, and sex. Only 0.03 per cent of records had differing data in three or more of these variables. A high proportion of the discordance was in the postcode which can legitimately change if the mother changes address between the babies’ birth and registration; it is also probably the variable most subject to recording and reporting errors.

Table 8 Gestational age in stillbirth registrations linked with NHS Number for Babies records

Gestational age (weeks)	Birth registration data																					Not stated	Total		
	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44			45	
24	57	4	1	1	.	1	.	.	.	1	65	
25	4	46	3	.	1	1	54	
26	.	5	38	3	.	1	.	1	48	
27	.	.	9	26	.	1	1	37
28	1	.	.	2	26	3	1	33
29	.	.	.	1	1	23	3	1	29
30	.	.	.	1	1	2	28	2	34
31	1	26	3	1	31
32	1	2	2	19	1	25
33	3	23	3	29
34	1	.	.	.	1	2	30	5	39
35	2	4	22	3	2	4	37
36	1	29	2	1	33
37	1	1	51	3	.	1	57
38	.	.	1	4	47	3	3	58
39	1	3	41	1	46
40	6	47	2	1	.	.	.	1	.	57
41	1	1	7	30	1	1	41
42	1	.	.	.	4	5
43	1	1	2
44	1	1
Total	63	55	52	34	30	30	35	30	27	27	38	30	33	62	54	51	57	33	7	.	.	.	1	12	761

Discrepancy in the recording of live/stillbirth status was found in less than 1 in 1,000 linked records. These rates of discordance, although lower than those observed in previous studies linking birth registration data with congenital anomaly notifications¹³ and with data in the maternity tail data of the Hospital Episode Statistics,¹⁴ clearly require further investigation. Classification at registration is judged to be the more reliable as a certificate of stillbirth is required to register a stillbirth and registrars should make enquiries before issuing a birth certificate if birth notification details sent from the child health department records the baby as stillborn. Errors made by maternity units in data submitted to NN4B Central Issue System cannot at present be amended. It can be difficult to classify babies who are born with ambiguous life signs and do not survive active resuscitation efforts. In some of these cases live or stillborn status could have been reclassified after sending data to NN4B.

Discrepancy in multiple birth status is found in approximately 1 in 1,000 linked records; as with live/stillbirth status, this is lower than observed in previous linkage studies.^{13,14} In addition to data recording errors there is a difference in definitions. NN4B data asks for the number of babies in this confinement and does not make a distinction between a registerable and a non-registerable birth. At registration multiple birth is flagged if there are two or more babies from one confinement registered. For example, in a twin pregnancy where the delivery is before 24 weeks gestation with one twin born dead and the other alive, only the latter is registered and the dead twin is classified as a late fetal death. Hence, the registration data would indicate only one birth was registered and would not flag it as a multiple. In contrast, the NN4B record would indicate that there were two babies in the confinement.

It would appear that the NHS number has been inaccurately recorded in some records during the process of data transfer from maternity units to registrars. This is evidenced by the finding of 11 registration records with NHS numbers issued to another baby, serious doubt about the accuracy of linkage using NHS numbers in 46 records, and the failure to find 114 of the NHS numbers in birth registrations amongst those issued by NN4B to babies born during that period. NN4B data come from the computer system which issues the NHS number and this number should, by definition, be correct. Further investigation of this is needed especially given the current NHS initiatives to develop systems that use NHS numbers for large-scale record linkage.¹⁵ Mechanisms to ensure that the NHS number is always electronically transferred would help eliminate mismatching due to errors in transcription. The process and quality of linkage could be improved substantially if the NHS number was recorded by ONS for stillbirths.

From the outset, the number of NN4B records exceeded the number of birth registration records for babies born in the same period. The extract of registration records for this project was taken from a provisional dataset to which further birth registrations have since been added. It is also possible that some births that occurred in private hospitals to non-resident foreign nationals were not registered. Minor differences in the number of NN4B and registration records may also occur due to discrepant recording of date of birth for births on the first and last days of any period reported. The possibility that a second NN4B record is generated for the same baby needs to be considered.

Birthweight was missing from the NN4B data in 0.4 per cent of the linked records. This was twice the percentage for which it was missing from the registration data. In the registration data there were more babies weighing 500–2,499g, and considerably fewer weighing under 500g, compared with the NN4B data. These findings are similar to those found in previous studies linking birth registrations with NHS records.^{13,14} Some babies, especially those who are very small or ill, may not be weighed immediately after birth and the weight may therefore not be available at the time the birth is notified to the NHS Numbers system. Missing or unusual birthweights, particularly those under 500g for live births, are

investigated by ONS by checking with registrars and also with child health departments. The NHS Numbers for Babies data provide the valuable opportunity to check birthweight information against gestational age, where neither data are missing. This will not, however, identify whether the recording error is in birthweight or gestational age.

Gestational age is not recorded at the registration of live births but, as it is recorded on the medical certificate of stillbirth, it is possible to compare the quality of gestational age data for stillbirths between registration and NN4B records. In this study, 82 per cent of linked stillbirth registrations had concordant data for gestational age. This is an improvement on the 64 per cent concordance found between gestational ages recorded for stillbirth registrations and maternity Hospital Episode Statistics records for births in 1997.¹⁴ Estimates of gestational age sent to obtain the NHS number are usually based on calculations based on the date of the woman's last menstrual period, supplemented in many cases by estimates from ultrasound scans. After a stillbirth occurs, details of the pregnancy are often reviewed and the gestational age may be recalculated. Examination of the baby's physical characteristics and x-ray studies may add further information which could alter the estimate of gestational age used to complete the medical certificate of stillbirth. It is possible that in some cases opinion about the gestation at which fetal death occurred and not the gestation at birth was reported.

This pilot work has shown a very good but not perfect degree of linkage. As a result of this pilot work, there may be some revision of the linkage process with the possibility of omitting some of the least productive stages. The next step will be to link data for the whole of 2005 using the final birth registration data. This will show which of our findings result from using a provisional registration extract and which persist even when the final extract is used. Once a full year's data are linked and their quality assessed we will be in a position to exploit the value added to birth registration data for England and Wales by the availability of additional items from this new data source, the NHS Number for Babies Service. It will become possible to enhance the statistics on infant and perinatal mortality, conceptions and fertility with information on the gestational age at live births. After further validation, other items notably data about the ethnicity of the baby and the more detailed information about place of birth recorded in NN4B records are also likely to prove useful.

Conclusions

This pilot study has shown that it is possible to link almost all NN4B records with a birth registration record. Although most records can be linked using NHS numbers we have shown that linkage using NHS numbers alone is not sufficient. This is the case for

- all stillbirth registrations (as they do not include NHS numbers)
- some live birth registrations that have no NHS number
- records that fail to link on the NHS number
- records that do link but where the link appears to be suspect because of discordant information in the two sources

If the NHS Numbers for Babies data, particularly the NHS number, were always transferred electronically to registrars for inclusion in the electronic registration record, and NHS numbers were retained in stillbirth registrations, this would facilitate the process and greatly improve the quality of linkage between NN4B and registration data.

There is a high level of agreement between the data held in the NN4B and registration sources, although this pilot study has thrown up some issues that need further investigation. This will include observing whether these problems persist once data for a full year are linked using the final birth registration data for 2005.

The NHS Numbers for Babies data are a reliable source of birth data which can be used to add gestational age and other data items to those available from birth registration for babies born in England and Wales.

Acknowledgements

This study would not have been possible without the generous help and support from the NHS Numbers for Babies team during the initial stages of obtaining the data. Collaborators in the National Gestational Age project included in addition to the authors, Lesz Lancucki, Maternity Hospital Episode Statistics, Community Health Statistics and Surveys, Health and Social Care Information Centre; Tony Couch, Head of Information Products, Health Solutions Wales; Gwyneth Thomas, Health Statistics and Analysis Unit, Welsh Assembly Government; Shona Golightly and Kate Fleming, Confidential Enquiries into Maternal and Child Health; and Martin Ward Platt, Clinical Director, Regional Maternity Surveys Office, North East Region.

References

1. Brocklehurst P (1999) Infection and preterm delivery. *British Medical Journal* **318**, 548–549.
2. ISD Scotland and Scottish Programme for Clinical Effectiveness in Reproductive Health (2004) *Scottish Perinatal and Infant Mortality and Morbidity Report 2003*, SPERCH Publication No 21, NHS Scotland: Edinburgh.
3. Confidential Enquiry into Maternal and Child Health (2004) *Stillbirth, neonatal and postneonatal mortality 2000–02, England, Wales and Northern Ireland*, CEMACH: London.
4. Department of Health (1974) NHS Reorganisation Circular HRC(74)3.
5. Hilder L and Alberman E (1998) Monitoring fetal and infant survival using regional birth notification data in north east London. *Journal of Epidemiology and Community Health* **52**, 253–258.
6. Stewart J H, Andrews J and Cartlidge P H T (1998) Numbers of deaths related to intrapartum asphyxia and timing of birth in all Wales perinatal survey, 1993–5. *British Medical Journal* **316**, 657–660.
7. Office for National Statistics (2006) *Mortality statistics: childhood, infant and perinatal. England and Wales, 2004*. Series DH3 number 37, ONS: London.
8. Devis T (2000) Recording of births and deaths in the countries of the United Kingdom. *Health Statistics Quarterly* **06**, 32–39.
9. National Health Service Act 1977, S124 (4–7).
10. Frith S and Jackson-Baker A (2002) The NHS Numbers for Babies (NN4B) Project. *Royal College of Midwives Journal* **5:4**, 120.
11. NHS Numbers for Babies website [cited August 2006] at: www.connectingforhealth.nhs.uk/nhsnumber/nn4b/
12. Statistical Analysis System (SAS). Version 8.2.
13. Botting B and Abrahams C (2000) Linking congenital anomaly and birth records. *Health Statistics Quarterly* **08**, 36–40.
14. Abrahams C and Davy K. (2002) Linking HES maternity records with ONS birth records. *Health Statistics Quarterly* **13**, 22–30.
15. Connecting for Health website [cited August 2006] at: www.connectingforhealth.nhs.uk/