

Linking maternity data for England, 2005–06: methods and data quality

Nirupa Dattani, Preeti Datta-Nemdharry and Alison Macfarlane *City University London, Department of Midwifery and Child Health.*

Abstract

Introduction

Maternity Hospital Episode Statistics (HES) data were linked to birth registration and NHS Numbers for Babies (NN4B) data to bring together some key demographic and clinical data items not otherwise available at a national level. This project added to earlier work involving linkage of birth registration records to NN4B records.

Methods

Birth registration and NN4B records were linked to Maternity HES delivery records and also Maternity HES baby records using the NHS Number or other indirect identifiers if NHS Number was missing.

Data quality and completeness of Maternity HES were assessed in relation to birth registration data wherever possible. For information not collected at registration, NN4B data were used to validate quality of Maternity HES.

Results

Overall, 91 per cent of Maternity HES delivery records could be linked to the birth registration/NHS Numbers for Babies records and 84 per cent of Maternity HES baby records were linked.

In 2005 only 3 per cent of Maternity HES records had mother's NHS number missing, compared with 30 per cent in the NN4B dataset. This did not reflect the extent to which Maternity HES data items were missing or discordant.

Over a quarter of all linked Maternity HES records for singleton babies had one or more of the following data items missing: birthweight, gestational age, birth status, sex, and date of birth of the baby. On the other hand, for data items where information was stated such as birthweight, birth status, and sex for singleton babies, there was good agreement between Maternity HES and linked birth registration and NN4B data.

Although NN4B records the ethnic category of the baby as defined by the mother, and Maternity HES records mother's ethnic category, 87 per cent of the linked records had the same ethnic group.

Conclusion

Even though a good linkage rate was obtained, the method used will be simplified before data for 2007 are linked. To gain the maximum benefit from this linkage in future years, improvements are urgently needed in the quality and completeness of the data contained in Maternity HES.

Contents

Abstract.....	1
Introduction	4
Methods	5
Results	9
Discussion.....	21
Conclusion	23
Acknowledgement.....	23
References.....	24
Appendix	25

List of Tables

Table 1	Number of linked records for singleton births with missing data items in common data fields, 2005	11
Table 2	Number of linked records for singleton births with missing data items in common data fields, 2006	11
Table 3	Comparison of plurality between registration and maternity HES, 2005 and 2006..	13
Table 4	Comparison of birth status for singleton births between registration and maternity HES, 2005 and 2006	14
Table 5	Comparison of sex for singleton births between registration and maternity HES, 2005 and 2006.....	15
Table 6	Comparison of birthweight distribution for singleton births between registration and maternity HES, 2005.....	16
Table 7	Comparison of gestational age for all births between NN4B and maternity HES, 2005.....	18
Table 8	Comparison of baby's ethnicity from NN4B with mother's ethnicity from maternity HES for all births, 2005.....	20
Table A1	Number and percentage of registration/NN4B records that were linked to HES mothers records by algorithm, 2005 and 2006	26
Table A2	Number and percentage of registration/NN4B records that were linked to HES baby records by algorithm, 2005 and 2006.....	27

Introduction

The data recorded at birth registration are mainly socio-demographic such as names, address of residence, place of birth, occupation of the parents and date of birth (shown in Office for National Statistics publication series DH3). As a result some important items needed for demographic and clinical purposes are not available at a national level. The opportunity to obtain gestational age and ethnicity nationally was provided by the introduction of the NHS Numbers for Babies (NN4B) Service in 2002. This service collects a small dataset which contains key items which are not recorded at birth registration. Information on gestational age at birth is of key importance. Babies born preterm, before 37 completed weeks of gestation, are at particular risk of morbidity and mortality in early years of life (Brocklehurst P, 1999; ISD Scotland report 2004; Confidential Enquiry into Maternal and Child Health, 2004)

Clinical information on maternity care at delivery could only be obtained from the Maternity Hospital Episode Statistics (HES) dataset for births that occurred in England and from the Community Child Health database (CHD) and Patient Episode Database for Wales (PEDW) for births that occurred in Wales.

Therefore a collaborative project between City University London, Office for National Statistics (ONS) and the Welsh Assembly Government was set up in 2004 to link these datasets for all births that occurred in England and Wales from 2005 to 2007. Stage 1 of the project involved linkage of birth registration data with NN4B dataset and assessment of data quality and completeness of the NN4B data. This is reported elsewhere (Hilder *et al.*, 2007 and Moser K *et al.*, 2008).

Stage 2 of the project, involving linkage of the dataset created in stage 1 to Maternity HES and the corresponding Welsh records for 2005 and 2006, and then the assessment of data quality and completeness by comparison with birth registration or NN4B, where possible. At the time, 2007 birth registration – NN4B linked data were not available. Therefore these data will be linked to Maternity HES and corresponding Welsh records at a later date using the experience gained in linking the first two years' data. This article describes linkage to Maternity HES records and reports on its data quality and completeness. The Welsh linkage for the first two years will be reported later.

Several data items are common to all three sources (Maternity HES, birth registration and NHS Numbers for Babies) as shown in Box 1. In addition, some data items are unique to each data source and linkage will enable new analyses across these data sources. For example, it will be possible to analyse caesarean section rates by father's socio-economic classification, compare time of birth with birth outcomes, and report on time and method of delivery by day of the week. Once the linkage has been completed and checked, the next stage of the project will be to undertake some of these analyses.

Box 1 Availability of selected data items from birth registration, NN4B and Maternity HES

Data items	Data sources		
	Birth registration	NN4B	Maternity HES
Baby's NHS number	+	+	+
Mother's NHS number		+	+
Birth date of baby	+	+	+
Delivery time		+	
Birthweight	+	+	+
Gestational age (stillbirth)	+	+	+
Gestational age (live birth)		+	+
Sex of baby	+	+	+
Number of babies born	+	+	+
Live or stillbirth	+	+	+
Parity (all births)			+
Baby/mother's postcode of usual residence	+	+	+
Ethnic category of baby		+	
Ethnic category of mother			+
Country of birth of mother	+		
Country of birth of father	+		
Father's socio-economic status	+		
Type of delivery place	+	+	+
Mother's date of birth	+	+	+
Marital status of mother	+		
Method of delivery			+
Complications in pregnancy			+

Methods

Source data

Birth registration

It is a legal requirement to register all live births within 42 days 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 (Office for National Statistics publication series DH3). The information is obtained, usually from one or both parents, by the local registrar of births, marriages and deaths. The local child health department passes some information from the birth notification it receives from the midwives to the registrar to verify the birth. Since 1975 this has included the baby's birthweight and since 2002 the NHS number (National Health Service Act 1977).

The process for registration of stillbirths is similar to that for live births, 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 (NN4B) service

The NHS Numbers for Babies Service was set up to issue an NHS number to all babies soon after birth: it went live in 2002. Under this system a standardised set of information about each birth is notified electronically, usually by the midwife attending the birth, to the Central Issuing System which checks for duplicates. If there is no existing record with the same details, the system returns a newly generated NHS number to the 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 notifier is required to check this record against the details of the baby they have just entered and use the existing number if a match is found. If no match is found then the notifier will proceed to obtain the NHS number using the NN4B Child Health Browser, or remotely by telephone from the NHS Number Issue Helpdesk if the browser is unavailable.

Records are held for six months on the Central Issuing System and then deleted. When the NHS number is issued, a copy of the information is sent to the local child health system where the birth occurred and a limited dataset, including the NHS number, to the NHS Central Register.

In the first part of the project, arrangements were made for ONS to receive a subset of variables from the NN4B dataset from 2005 onwards for linkage to birth registration record.

Maternity Hospital Episode Statistics (HES)

Maternity Hospital Episode Statistics contains data for all births occurring in England, including those at home and in non-NHS hospitals. There are however, some data items not recorded for births outside NHS hospitals such as admission/discharge dates or patient classification. HES includes a wide range of information on maternity such as details on how the baby was born (method of delivery), method of onset of labour, complications, gestation and ethnicity as well as information about the baby, such as date of birth, gender, birthweight and geographical information on where the baby was born.

Most records of admission to hospital will be classed as general inpatient episodes, even when heavily pregnant women are admitted to maternity wards with the clear intention of giving birth. However, as soon as the mother has given birth, the record becomes a maternity record and is updated and extended before being submitted to HES. If the pregnant woman does not give birth, for whatever reason, before the episode ends, the record remains as a general inpatient record (known as a general episode in HES).

There are two types of maternity records in HES, the delivery record and the birth record. Both types of records consist of an admitted patient care record with an additional 19 fields, in an appended baby 'tail'.

- A delivery record is a HES record for the mother containing the same data as a general record, but has a baby tail for information about the delivery
- A birth record is a HES record for the baby. Again it has the same format as a general record but it also has a baby tail containing exactly the same information recorded in the corresponding tail of the delivery record

The baby tail data coverage is not as complete as the rest of the HES data (HES website). There are a number of reasons for the incompleteness and data quality issues, such as:

- Trusts submitting a significantly higher number of delivery episodes compared to birth episodes
- Trusts failing to submit data on the number of birth episodes where they record a high number of delivery episodes
- Trusts failing to submit delivery records – the reason for this is that approximately 20 trusts have a stand-alone maternity system which is not linked to the Patient Administration System
- Trusts reporting a high number of maternity beds available, but not recording any information about deliveries or births
- Trusts reporting that they have no maternity beds available, but recording a high number of birth and delivery episodes

Over the years from 2001–02 to 2005–06, coverage of deliveries was 86.2 per cent on average. The problems with coverage are significantly compounded by data quality issues (HES website)

HES Patient ID (HESID) is an identifier used to track patients through the HES database, or for linkage to other datasets such as mortality, rather than using patient identifiers such as NHS Number. It is a pseudonymised number which uniquely identifies each patient without the necessity of viewing or using patient identifiable information such as the NHS Number. HESID is derived using a matching algorithm which looks at various combinations of the following patient identifiable fields:

- NHS Number (fieldname NEWNHSNO)
- Date of Birth (fieldname DOB)
- Sex (fieldname SEX)
- Postcode (fieldname HOMEADD)
- Provider code (fieldname PROCODET)
- Local patient ID (fieldname LOPATID)

HESIDs are stored in the HES index. This is updated monthly and older versions are not kept. Hence the most recent HES index (22 June 2010) was used to link birth registration/NHS Numbers for Babies data for 2005 and 2006.

For each episode of care with a particular consultant or midwife, a HES record is created, but each time this record is updated with new information, a new version of the record is created. As a result, several versions of the record of the same episode are created.

Record linkage

Record linkage was carried out by Northgate Solutions, which processes HES records under contract with the NHS Information Centre. For details on linkage of registration and NN4B linked data to the mother and baby records in HES, see Appendix.

The linked data for each year (2005 and 2006) consisted of previously linked registration and NN4B data linked to the mother record in HES, and a second file based on linkage of previously linked registration and NN4B data to the baby record in HES. These were provided to ONS and were accessed by researchers from City University London in the secure environment of the Virtual Microdata Laboratory (VML) facilities at ONS. Outputs of analyses undertaken in the VML were released by ONS in the form of disclosure controlled tables.

Data Quality

Review of quality of Maternity HES was based on completeness and consistency of the HES data in relation to birth registration data wherever possible. Since all babies born in England and Wales have to be registered, information collected at registration is subject to quality checks (Office for National Statistics publication series DH3). However, where information was not available from registration, NN4B data were used to validate the quality of Maternity HES. The quality of the NN4B data in comparison to birth registration data is reported elsewhere (Moser *et al.*, 2008). Completeness of the main data items in all three sources was measured by identifying missing data.

The linked data for the mothers file was split into singleton and multiple births (using multiple birth status information from registration) to ease assessment of data quality. In some instances the results are reported separately.

Results for 2005 are reported in this paper and 2006 results are available on the [ONS website](#) in a spreadsheet

Data analyses were carried out using SAS version 9 and SPSS version 16.

Results

Mother file

The Maternity HES record is a mother-based record containing the mother's details with a maternity tail and a baby tail which can accommodate up to nine babies born in one maternity. In contrast, the registration and NN4B linked data consists of one record per baby. Therefore, the linkage was based on baby to mother record.

Northgate solutions returned 582,680 records that had linked to 2005 registration and NN4B linked data. For 2006, 601,623 HES records were linked. These included multiple records for the same mother for each episode. Records with the most information were selected to ensure one-to-one linkage to registration and NN4B linked dataset. This gave a file of 566,240 records for 2005 and 584,795 records for 2006.

Due to the way the linkage was done, the same Maternity HES record was linked to multiple registration and NN4B linked records. Maternity HES records where the postcode, mother's or baby's date of birth and birthweight were all inconsistent, when compared to registration and NN4B linked dataset were deleted. A total of 264 and 364 Maternity HES links were deleted for 2005 and 2006 data respectively.

In the 2005 registration and NN4B linked data there were 617,613 babies who were either born in England or resident in England. The resident in England category was used for births that occurred at home in the registration and NN4B linked data.

Around 64 per cent of the linked registration and NN4B records were linked to Maternity HES records using mother's NHS number and partial date of birth.

A further 27 per cent of the linked registration and NN4B records were matched to Maternity HES using mother's postcode and date of birth. Registration and NN4B linked records that were not linked to HES accounted for 8 per cent of all records. Of the Maternity HES records, 3 per cent did not link to registration and NN4B linked records.

Overall, there were 50,380 singletons and 1,265 multiples in the registration and NN4B linked data that were not linked to HES records. This gave a total 565,968 Maternity HES records that were linked to the registration and NN4B linked records giving a linkage rate of 91.6 per cent. The linkage rate for 2006 was similar at 91.3 per cent.

Baby file

The baby file was much more straightforward than the mother file as it involved a one to one link between baby records in registration and NN4B linked data, and Maternity HES.

For 2005, a total of 686,087 HES baby records were linked to registration and NN4B linked data by Northgate solutions. These included 128,482 records that were general episodes. After omitting these, there were 557,605 HES baby records that had been linked to registration and NN4B linked data. There were multiple HES birth records for the same baby linked to registration and NN4B linked record. Again only the records with the most information were kept and others were deleted.

After deletion, 524,536 records remained. On further examination of the data, 4,009 records had to be deleted where the same HES record was matched to another registration and NN4B linked record. A similar process was applied for the 2006 data with 674,534 records sent by Northgate and 540,283 linked HES records were left after deletion of multiple HES baby record. A further 1,958 records were deleted due to same Maternity HES records being linked to more than one registration and NN4B linked record.

In the 2005 registration and NN4B linked data, there were 617,613 babies who were either born in England or resident in England; 515,434 registration and NN4B linked records were linked to HES baby records using NHS number, partial date of birth, and sex; and 4,597 were linked using date of birth, postcode, and sex. Over 15 per cent of registration and NN4B linked records were not linked to HES baby records. Around 1 per cent of HES baby records did not link to registration and NN4B linked data.

Overall 520,527 out of 617,613 records were linked giving a linkage rate of 84.3 per cent. For 2006 the linkage rate was 84.1 per cent.

Data Quality

For HES, missing and discordant data were assessed only in the mother record – as this included information on the baby, and also because the linkage rate was far better than the baby record. For multiple births information was recorded only on the first baby; data on other babies was either missing or the same as the first baby suggesting there were problems in the linkage process in HES. Hence singleton and multiple births were analysed separately and only results for singletons are reported here.

Missing data

The mother's NHS number is recorded only on the NN4B record and not recorded at birth registration. For singleton births, 30 per cent of registration and NN4B linked records did not have the mother's NHS number compared with three per cent in the Maternity HES records. In Maternity HES, birthweight and gestational age information was missing for 25 per cent and 48 per cent of singletons respectively. Status of baby, date of birth of baby and sex of baby was missing in over 25 per cent of the records (Table 1). Similar results were observed in 2006 (Table 2).

Table 1 **Number of linked records¹ for singleton births with missing data items in common data fields, 2005**

England

	NHS Numbers for Babies		Registration		Maternity HES	
	Number	Percent	Number	Percent	Number	Percent
NHS number of mother	164,458	30.0	NA	NA	16,685	3.0
Date of birth of mother	960	0.2	0	0.0	0	0.0
Ethnicity	59,865	10.9	NA	NA	77,771	14.2
Postcode	10,101	1.8	10	<0.1	693	0.1
Birthweight	2,721	0.5	874	0.2	135,144	24.6
Gestational age	3,829	0.7	NA	NA	264,877	48.3
Status	615	0.1	0	0.0	176,455	32.2
Date of birth of baby	615	0.1	0	0.0	139,414	25.4
Sex of baby ²	1,098	0.2	0	0.0	144,115	26.3

1 Number of linked records 548,857

2 Includes 483 cases with indeterminate sex in NN4B and 15 cases in maternity HES

Source: HES, registration and NHS Numbers for Babies

Table 2 **Number of linked records¹ for singleton births with missing data items in common data fields, 2006**

England

	NHS Numbers for Babies		Registration		Maternity HES	
	Number	Percent	Number	Percent	Number	Percent
NHS number of mother	147,782	26.1	NA	NA	14,440	2.6
Date of birth of mother	465	0.1	0	0.0	0	0.0
Ethnicity	56,521	10.0	NA	NA	90,839	16.0
Postcode	8,209	1.5	14	<0.1	343	0.1
Birthweight	3,123	0.6	2,839	0.5	142,696	25.2
Gestational age	4,104	0.7	NA	NA	255,437	45.1
Status	189	0.0	0	0.0	155,754	27.5
Date of birth of baby	189	0.0	0	0.0	146,183	25.8
Sex of baby ²	704	0.1	0	0.0	153,015	27.0

1 Number of linked records 566,313

2 Includes 515 cases of indeterminate sex in NN4B and 1,529 cases in Maternity HES

Source: HES, registration and NHS Numbers for Babies

Discordant data

Discordance in common individual data items

Discordance in each of the common data fields in the linked records was assessed using information from birth registration rather than NN4B. However, for data items that were not recorded at birth registration, NN4B data were used.

For singletons, using birth registration and Maternity HES data, baby's date of birth was discordant in 0.2 per cent of records in 2005 and 2006. Postcode did not agree in 7 per cent of records in 2005 and 2006.

Discordance in multiple birth status

There were 13,850 records that were identified as multiple births in birth registration and Maternity HES. Multiple birth status was discordant between the two data sources in 1,806 records in 2005 and 1,913 records in 2006 (Table 3). In 2005, 91,030 Maternity HES records had unknown multiple birth status.

Table 3 **Comparison of plurality¹ between registration and maternity HES, 2005 and 2006**

England

Year	Maternity HES plurality	Registration plurality		
		Singleton	Multiple	Total
2005	Singleton	459,282	178	459,460
	Multiple	1,628	13,850	15,478
	Total	460,910	14,028	474,938
2006	Singleton	461,912	267	462,179
	Multiple	1,646	14,543	16,189
	Total	463,558	14,810	478,368

¹ Plurality was missing in maternity HES for 91,030 records in 2005 and 106,051 records in 2006.

Source: HES and registration

Discordance in live or stillbirth status

For the records which had a stated live or still birth status in both data sources, 0.05 per cent and 0.2 per cent did not agree in 2005 and 2006 (Table 4) respectively. Out of all the linked records for 2005, around 32 per cent of Maternity HES records did not have information on birth status. This fell to 27.5 per cent in 2006.

Table 4 **Comparison of birth status for singleton births between registration and maternity HES, 2005 and 2006**

England

Year	Maternity HES birth status	Registration birth status			
		Live birth Number	Still birth Number	Total Number	Percent
2005	Live	370,939	99	371,038	67.6
	Still birth: ante-partum	56	1,021	1,077	0.2
	Still birth: intra-partum	9	137	146	0.0
	Stillbirth: indeterminate	6	135	141	0.0
	Not Known	175,578	877	176,455	32.2
	Total	546,588	2,269	548,857	
2006	Live	408,178	114	408,292	72.1
	Still birth: ante-partum	63	1,212	1,275	0.2
	Still birth: intra-partum	14	178	192	0.0
	Stillbirth: indeterminate	661	139	800	0.1
	Not Known	154,883	871	155,754	27.5
	Total	563,799	2,514	566,313	

Source: HES and registration

Discordance in baby's sex

The sex of the baby recorded on birth registration for singleton births was used to compare with Maternity HES. In 2005 where baby's sex was stated in both data sources, an agreement of 99 per cent was observed (Table 5). A similar percentage was noted in the 2006 data. In the 2006 Maternity HES data, sex was coded incorrectly in 301 records to codes 4, 5, 6, and 7, and there were 1,529 records where sex was coded as indeterminate, compared with 15 records with indeterminate sex in 2005.

Table 5 **Comparison of sex for singleton births between registration and maternity HES, 2005 and 2006**

England

Year			Registration			Percentage
			Male	Female	Total	
2005	Maternity HES ¹	Male	204,613	791	205,404	50.8
		Female	2,814	196,524	199,338	49.3
		Total	207,427	197,315	404,742	
2006	Maternity HES ²	Male	208,162	586	208,748	50.7
		Female	2,978	200,043	203,021	49.3
		Total	211,140	200,629	411,769	

1 There were 34 records classified as not known, 78 as not specified, 15 as indeterminate, 143988 records had sex missing.

2 There were 2323 records classified as not known, 1450 as not specified, 1529 as indeterminate, 148941 records had sex missing and 301 records had sex coded incorrectly.

Source: HES and registration

Discordance in birthweight

Birthweights in birth registration data were grouped into 500g groups and compared with grouped birthweights from Maternity HES. In terms of concordance between the two data sources, 99 per cent of records with stated birthweight were in the same 500g birthweight group. This amounts to only 75 per cent of all linked records because birthweight was not stated on a large number of records. In Maternity HES, birthweight was missing in 25 per cent of the records compared with only 0.2 per cent in birth registration data (Table 6). Similar findings were also observed with the 2006 data (99 per cent concordance between the two data sources where birthweight was stated) and these can be found in [Table 1 on the web](#).

Table 6 **Comparison of birthweight distribution for singleton births between registration and maternity HES, 2005**

England

Registration Birthweight (grams)	Maternity HES birthweight (grams)												Not stated	Total	Percentage
	<500	500–999	1000–1499	1500–1999	2000–2499	2500–2999	3000–3499	3500–3999	4000–4499	4500–4999	5000–5499	5500+			
<500	233	4	4	5	4	3	10	5	10	5	1	.	140	424	0.1
500–999	5	1,533	.	7	7	9	4	4	4	2	2	17	693	2,287	0.4
1000–1499	7	6	2,322	11	2	7	17	7	5	1	.	.	954	3,339	0.6
1500–1999	11	3	12	4,638	6	21	12	7	2	.	.	.	1,716	6,428	1.2
2000–2499	18	3	7	17	16,259	250	308	174	65	10	1	3	5,408	22,523	4.1
2500–2999	30	7	6	15	112	68,622	162	236	22	6	6	14	21,888	91,126	16.6
3000–3499	93	14	31	16	121	160	149,372	558	100	5	17	24	48,764	199,275	36.3
3500–3999	62	10	13	11	30	205	508	119,583	97	13	8	21	39,781	160,342	29.2
4000–4499	30	1	11	3	10	16	69	64	39,334	39	6	7	13,051	52,641	9.6
4500–4999	8	.	.	2	1	2	13	29	42	6,387	.	.	2,155	8,639	1.6
5000–5499	.	1	.	1	.	1	6	5	1	.	636	.	216	867	0.2
5500 +	1	7	1	2	2	3	1	2	.	.	.	47	26	92	0.0
Not stated	6	14	5	7	32	93	203	117	39	5	1	.	352	874	0.2
Total	504	1,603	2,412	4,735	16,586	69,392	150,685	120,791	39,721	6,473	678	133	135,144	548,857	
Percentage	0.1	0.3	0.4	0.9	3.0	12.6	27.5	22.0	7.2	1.2	0.1	0.0	24.6	100.0	

Source: HES and registration

Discordance in gestational age

Information about gestational age for all births was available from the NN4B and Maternity HES. In 2005 approximately 89 per cent of records with a recorded gestational age were the same in both data sources (Table 7). For Maternity HES data 48 per cent of records had gestational age information missing. Gestational age differed by one week in around 7 per cent of the records, and two weeks or more in about 4 per cent of the records. There was a large variation in gestational age between the two data sources in the 'tails' for those under 30 weeks and over 42 weeks, but only 3 per cent of births occurred in these gestational age groups. The difference ranged from 20 to 85 per cent for those under 30 weeks. At 43 weeks, gestational age differed in about half of the records and it decreased further to a third of all records at 44 weeks and over.

For 2006 there was agreement between the two sources in 90 per cent of records which contained gestational age from Maternity HES and NN4B ([Table 2 on the web](#)).

Table 7 Comparison of gestational age for all births between NN4B and maternity HES, 2005

England

NN4B Gestational age (weeks)	Maternity HES gestational age (weeks)																								Not stated	Total
	<22	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44 +		
<22	74	.	.	.	2	1	1	.	2	1	1	2	.	1	2	7	3	.	.	.	121	218
22	2	44	1	1	1	74	123
23	7	2	89	7	2	1	.	1	.	.	.	114	223
24	6	.	5	235	8	3	.	1	.	2	.	2	1	.	.	.	323	586
25	12	.	.	6	253	9	3	.	3	2	2	2	1	.	.	269	562
26	9	1	1	1	6	312	11	1	1	.	1	.	1	1	.	4	.	1	1	.	1	.	.	.	373	726
27	11	.	.	3	.	7	356	10	.	.	1	2	.	3	3	2	2	.	.	374	774
28	10	1	.	.	1	5	9	465	9	.	1	.	.	2	1	1	2	11	2	2	4	1	.	.	521	1,048
29	5	.	.	.	2	.	.	11	470	21	1	1	4	1	.	.	3	.	5	5	626	1,155
30	12	1	1	15	611	16	6	6	1	1	.	1	3	7	14	788	1,483
31	14	.	.	1	.	.	2	.	3	18	770	25	.	3	3	.	.	3	6	17	3	.	.	.	910	1,778
32	1	17	.	1	.	.	.	2	.	4	20	1,168	21	9	1	6	2	5	1	24	3	.	.	1	1,241	2,527
33	1	.	36	1	1	1	9	33	1,529	54	10	8	9	8	7	16	2	.	.	1	1,824	3,550
34	.	.	45	3	2	2	.	1	1	2	6	3	57	2,474	85	19	19	18	14	37	3	.	.	.	2,814	5,605
35	3	.	.	58	1	2	5	2	14	72	3,826	125	27	17	16	63	16	2	.	1	4,119	8,369
36	4	.	1	.	102	.	1	3	1	1	2	5	6	24	132	7,125	227	62	33	141	20	11	1	9	7,616	15,527
37	5	.	.	.	229	1	6	2	2	.	3	7	5	7	26	269	15,049	611	162	312	67	28	7	13	15,697	32,508
38	7	3	.	1	2	473	2	4	3	11	6	11	24	23	28	80	751	36,057	1,405	881	114	50	26	36	37,091	77,089
39	6	.	2	.	.	2	881	1	3	10	7	9	9	18	32	68	169	1,610	55,910	2,915	406	123	74	83	59,341	121,679
40	13	.	3	1	4	1	3	983	1	5	3	9	5	21	25	70	98	331	2,444	71,064	2,593	368	147	180	72,955	151,327
41	7	.	.	2	2	2	1	865	1	4	9	4	7	11	14	44	63	121	307	4,411	49,742	1,275	192	228	52,850	110,162
42	2	1	.	.	.	1	.	.	252	.	.	2	2	5	2	6	24	54	81	548	1,290	8,452	62	49	11,205	22,038
43	1	1	.	12	2	.	.	.	1	2	7	22	51	102	74	42	543	7	582	1,449
44+	1	2	.	6	.	.	.	3	2	2	6	9	26	73	43	26	.	278	397	874
Not stated	.	1	.	3	3	5	5	8	8	11	10	14	15	36	53	115	239	458	679	780	616	88	1	4	1,436	4,588
Total	211	70	183	322	619	821	1,284	2,361	776	719	872	1,299	1,705	2,769	4,245	7,947	16,698	39,404	61,164	81,420	55,006	10,469	1,053	890	273,661	565,968

Source: HES and registration

Discordance in ethnicity

Baby's ethnicity from the NN4B dataset and mother's ethnicity from Maternity HES were compared (Table 8). There was agreement in 87 per cent of the records which had a stated ethnic category. Out of all the linked records, 19 per cent of records had no ethnicity recorded in Maternity HES, and in 11 per cent of records ethnic group was not stated in the NN4B dataset.

For 2006 data the ethnic category was the same in 87 per cent of the linked records. In Maternity HES 16 per cent of records had no ethnicity recorded and 10 per cent of records had ethnic group not stated in the NN4B dataset ([Table 3 on the web](#)).

Table 8 **Comparison of baby's ethnicity from NN4B with mother's ethnicity from maternity HES for all births, 2005**

England

Babies ethnicity	Mother's ethnicity																Total	
	White British	White Irish	Any other White	White and Black Caribbean	White and Black African	White and Asian	Any other mixed	Indian	Pakistani	Bangladeshi	Any other Asian	Caribbean African	Any other Black Chinese	Any other	Not stated			
White British	281,674	1,036	7,385	179	75	186	441	254	255	90	183	183	209	99	118	1,440	68,309	362,116
White Irish	683	832	76	2	.	1	6	.	1	.	3	3	3	.	3	16	306	1,935
Any other White	5,746	170	14,963	20	46	42	403	26	23	6	130	17	75	36	31	1,607	4,262	27,603
White and Black Caribbean	2,577	24	134	558	34	5	110	1	2	1	4	503	79	151	.	146	853	5,182
White and Black African	1,003	13	144	21	278	1	42	6	3	2	9	49	700	170	1	84	596	3,122
White and Asian	1,723	34	220	9	2	274	65	342	120	32	392	4	10	7	117	260	921	4,532
Any other mixed	2,259	32	645	175	67	81	615	180	87	31	249	168	134	150	204	729	1,351	7,157
Indian	290	.	36	23	3	39	36	9,553	855	220	872	18	37	15	2	275	2,370	14,644
Pakistani	300	1	34	.	3	44	22	439	17,515	224	732	7	27	38	2	202	2,716	22,306
Bangladeshi	92	1	12	2	1	3	7	120	427	5,974	175	4	8	6	.	68	526	7,426
Any other Asian	132	1	76	9	6	57	65	647	361	104	2,721	18	52	32	125	951	1,298	6,655
Caribbean	283	6	38	218	25	1	53	5	6	2	15	4,195	274	592	1	147	788	6,649
African	234	1	107	28	541	3	67	34	28	18	97	292	11,585	1,111	1	500	2,376	17,023
Any other Black	182	3	71	74	81	4	47	24	11	5	51	390	591	2,278	2	181	666	4,661
Chinese	62	2	16	.	1	5	24	2	.	1	68	.	2	.	1,323	135	380	2,021
Any other	1,198	26	1,249	80	62	77	255	163	130	35	837	92	378	163	149	4,032	2,233	11,159
Not stated	35,701	197	2,058	185	121	118	267	2,076	915	698	498	326	1,097	316	263	1,192	15,749	61,777
Total	334,139	2,379	27,264	1,583	1,346	941	2,525	13,872	20,739	7,443	7,036	6,269	15,261	5,164	2,342	11,965	105,700	565,968

Source: HES and registration

Discussion

Northgate Solutions designed part of the linkage algorithm based on its previous experience of linkage and this was enhanced by the authors to improve the linkage rate. Northgate also linked Maternity HES records (mother and baby records) to registration and NN4B linked data. There were however, some issues with the linked file as in some cases more than one registration and NN4B linked records were linked to the same Maternity HES record using different combinations of patient identifiable information. Also in the mother file, delivery records were provided (as requested) but in the baby file both birth and general episodes were provided when only the former were requested. Furthermore, HES data included multiple records for each episode. These had to be reviewed to identify the record with the most information that should be kept for analysis and the rest were deleted. Hence, a considerable amount of time was spent in cleaning the files. The final files consisted of one maternity record linked to a registration and NN4B linked record for the mother, and one HES baby record linked to a registration and NN4B linked record for the baby. The linkage method used by Northgate Solutions will be simplified before data for 2007 are linked, so that a maternity HES record is linked to a single linked registration and NN4B record.

Two-thirds of the registration and NN4B records were linked to the HES mother record using NHS number and partial date of birth. This was not surprising as mother's NHS number was missing in nearly a third of the registration and NN4B linked records and also a very small proportion of Maternity HES records. A further quarter of the registration and NN4B linked records were linked using date of birth or month and year or birth, and postcode. There were concerns about using postcode in the linkage algorithm, as the HES index used for linkage is derived using current postcode of residence of the mother and the postcode on registration and NN4B linked data was recorded at the time of registration. It is possible the mother could have moved since having the baby and this variable is also subject to recording and reporting errors. Despite this, an overall linkage rate of over 90 per cent was achieved for both 2005 and 2006. This could have improved further if there was a shorter delay before linkage was carried out, as HESID would be less likely to have changed, or if HESID at birth was retained as a separate field for linkage.

The linkage rate for registration and NN4B linked data to HES baby record was slightly lower than the linkage rate for the mother record. This was not surprising, as large proportion of 'baby tails' are known to be missing in Maternity HES ([see HES website](#))

HES mother records include information on the baby, and as the linkage rate for registration and NN4B linked data to HES mother records was better than the baby records quality of information on HES was assessed using the mother records. There were however, issues with multiple births in the HES mother record. Information was often recorded for the first baby only and for the remaining multiples it was either missing or the same as the first baby. But in very few cases it was different. Multiple birth status was also unknown in a fifth of the records. Further work to assess quality of data on multiple births is necessary before it could be used for any analyses.

Discrepancy in the recording of live/still birth status for singleton babies was found in 5 in 10,000 linked records in 2005, and 2 in 1,000 records in 2006. Classification at birth registration is judged to be more reliable as a medical certificate of stillbirth is required to register a stillbirth. Also a third of the HES records did not have any information on birth status; this is a much higher proportion

than found in pilot study involving linkage of Maternity HES data for one month to birth registration (Abraham C *et al.*, 2002).

Birthweight was missing in a quarter of all linked Maternity HES records for singletons babies, compared with only 0.2 per cent at birth registration. There was however, good concordance between the two data sources where birthweight was stated, as 99 per cent of records were in the same 500g birthweight group. Missing birthweights are investigated by ONS by going back to registrars and also to child health departments. Therefore the quality of birthweight information on registration is expected to be better and more reliable than in Maternity HES.

Gestational age is not recorded at registration for live births but is available from the NN4B data. This records gestational age in weeks 'calculated from relevant menstrual data held within the maternity system' whereas Maternity HES specifies 'time from the first day of the last menstrual period (LMP)'. Where this is not available an estimate is supposed to be recorded. However, it is likely the gestational age assessed by ultrasound is now used because second trimester scans are a routine part of antenatal assessment in Britain. A study of births at 27/28 weeks of gestational age in England, Wales, and Northern Ireland between 1998 and 2000 showed that 79 per cent of the mothers had had an ultrasound before 20 weeks gestation, and 85 per cent had had menstrual history recorded (Confidential Enquiry into Stillbirths and Deaths in Infancy, 2001). Gestational age distributions have been shown to differ depending on the method used to assess gestational age. Studies have shown that if second trimester ultrasound is used rather than LMP, then the mean gestational age is one week less. However, gestational age differed by one week in only 7 per cent of the linked records. Of all the linked records, nearly half of the HES records had no information on gestational age, and where gestational age was stated it was in good agreement with NN4B in majority of the records.

The NN4B system requests information on the ethnic category of the baby, as defined by the mother using the 2001 Census categories (Moser K *et al.*, 2008). On the Maternity HES record it is the mother's ethnicity which is self-selected using the 2001 Census categories. It is however, unclear in both of these data sources, whether the mother was involved in defining the ethnic category or the health professional decided what to record rather than asking the mother. In practice it is likely to be a mixture of both, and although the ethnic category of the baby is requested in NN4B it is not possible to know whose ethnicity was actually recorded, the mother's or the baby's. A further consideration is that people's identification with an ethnic group is not always straightforward and individual responses, whether self-reported or not, may vary according to circumstances and over time. Despite these limitations, in over 80 per cent of the linked records mother's ethnicity was the same as that of the baby. In 4.2 per cent of records mother's ethnicity has been categorised as 'White British' and baby's ethnicity has been categorised as 'White other' or vice versa. This suggests that father's ethnicity may have been taken into consideration in recording the baby's ethnic category on the NN4B data and this is more likely to have been defined by the mother.

Conclusion

This study shows that it is possible to link the majority (90 per cent) of Maternity HES records to registration and NN4B linked records, but the method used for linkage by Northgate Solutions needs to be amended before it is used to link data for future years. Linkage would be beneficial, and should be carried out routinely, if data quality and completeness improves in Maternity HES. However, at national level, information such as parity, method of delivery, complications in pregnancy is only available from Maternity HES, so linkage would be necessary to access this information, together with the data obtained from birth registration and NN4B.

Birth registration and NN4B are more reliable sources of data than Maternity HES. But where data have been recorded they are in good concordance with birth registration or NN4B, but there were a large proportion of linked records where information was not recorded on Maternity HES.

Acknowledgement

This work forms part of the linkage, analysis and dissemination of national birth and maternity data for England and Wales project, funded by the Medical Research Council as part of the Joint Wellcome Research Councils Electronic patient data linkage initiative. We would like to thank Northgate Solutions, in particular Jonathan Low, for linking the datasets; Julie Messer at ONS for providing the birth registration – NHS Numbers for Babies linked data to Northgate Solutions to link to Maternity HES records, making the linked data accessible in the VML system and for releasing outputs; Chris Roebuck and Tony Childs at the NHS Information Centre, for their advice and support. Collaborators in the original National Gestational Age project included, in addition to the authors, Lesz Lancucki, formerly Maternity Hospital Episode Statistics, Community Health Statistics and Surveys, NHS Information Centre and Tony Couch, formerly Head of Information Products, Health Solution Wales who we would like to thank for their help in the earlier stage of the project. We are grateful to Gwyneth Thomas, Health Statistics and Analysis Unit, Welsh Assembly Government and Martin Ward Platt, Clinical Director, Regional and Maternity Surveys Office, North East Region for their help and support in this project.

Maternity HES data are copyright © 2009, Re-used with the permission of The Health and Social Care Information Centre. All rights reserved.

References

- Abrahams C and Davy K (2002). Linking HES maternity records with ONS birth records. *Health Statistics Quarterly* 13, 22–30
- Brocklehurst P (1999). Infection and preterm delivery. *British Medical Journal*, 318, 548–549.
- Confidential Enquiry into Maternal and Child Health (2004) Stillbirth, neonatal and postneonatal mortality 2000–02, England, Wales and Northern Ireland.
- Confidential Enquiry into Stillbirths and Deaths in Infancy (2001) *8th Annual Report*, Maternal and Child Health Research Consortium: London.
- Hilder L, Moser K, Dattani N and Macfarlane A (2007). Pilot linkage of NHS Numbers for Babies data with Birth registrations. *Health Statistics Quarterly* 33,25–33.
- 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
- ‘Maternity data in HES’ available on the Information Centre website at:
www.hesonline.nhs.uk/Ease/servlet/ContentServer?siteID=1937&categoryID=925
- Moser K and Hilder L (2008). Assessing quality of NHS Numbers for Babies data and providing gestational age statistics. *Health Statistics Quarterly* 37,15–23
- Moser K, Stanfield K M and Leon D A (2008). Birthweight and gestational age by ethnic group, England and Wales 2005: introducing new data on births. *Health Statistics Quarterly* 39,22–31.
- National Health Service Act 1977, S124 (4–7).
- Office for National Statistics, *Mortality Statistics: Childhood, infant and perinatal, England and Wales, 2007*. Series DH3 No. 40. Available on the ONS website
www.statistics.gov.uk/StatBase/Product.asp?vlnk=15362

Appendix

For linkage of registration and NHS Numbers for Babies (NN4B) linked data to the mothers records (also known as delivery record) in HES, a file consisting of a small subset of data items including the mother's NHS number, date of birth and age, the baby's date of birth, their postcode and a unique ID compiled by ONS was sent to Northgate Solutions. A second file consisting of baby's NHS number, date of birth, postcode of residence and the unique id was sent to link to the baby records, known as 'baby tails'. Northgate extracted the mother and baby records from HES using the filters shown in Box A1. These records were then linked to the registration and NN4B linked records.

Box A1 The filters used by Northgate Solutions to extract HES records for mother and baby

Mother records

Episode type = 2 or 5 (delivery episodes)

AND

Main procedure/Intervention Code between R14 and R27

OR

Main diagnosis code between ICD codes O00 and O99 (O = pregnancy and childbirth and Puerperium conditions).

Baby records

Episode type = 3 or 6 (birth episodes)

AND

Well baby index = YES

OR

Neonatal level of care in (0,1,2,3,4,9)

OR

Main diagnosis code between ICD codes P00 and P99 (Conditions originating in perinatal period)

OR

Main diagnosis code between ICD codes Q00 and Q99 (Congenital malformations, deformities, chromosomal abnormalities).

For the registration and NN4B linked data, indirect identifiers were used for linkage of records where the mother's NHS number was missing. These included different combinations of mother's date of birth, postcode, and baby's date of birth. A pilot study, using registration and NN4B linked data for all babies that were born in January 2005, was carried out to test the algorithm compiled by Northgate Solutions, involving stages 1–3 as shown in Table A1. This was based on their previous experience of linking mortality registration records to HES.

Examination of the unlinked registration and NN4B linked data and maternity HES records showed that the linkage rate could be improved by using partial information, such as month and year of birth or first four characters of postcode. Therefore further stages were added to the algorithms

used for linking the mother and baby records before the annual registration and NN4B linked data files for 2005 and 2006 were sent to Northgate Solutions. The final algorithm used to link mothers records to the registration and NN4B linked data involved 9 stages, of which linked records were obtained only from stages 1, 2 and 5 (Table A1). A variable indicating stage of algorithm used for linking HES records to the registration and NN4B linked dataset was provided by Northgate Solutions.

Table A1 **Number and percentage of registration/NN4B records that were linked to HES mothers records by algorithm, 2005¹ and 2006²**

England

Stage	Variables used	Records linked			
		2005		2006	
		Number	Percentage	Number	Percentage
1	Exact NHSNO + partial mother's DOB	395,855	64.1	431,296	67.4
2	Exact mother's DOB + exact postcode	168,704	27.3	152,040	23.8
3	Exact NHSNO + exact mother's DOB	-	-	-	-
4	Exact NHSNO + month and year of mother's DOB	-	-	-	-
5	Exact postcode + month and year of mother's DOB	1,409	0.2	1,083	0.2
6	Exact baby's DOB + baby sex + exact POSTCODE	-	-	-	-
7	Exact baby's DOB + exact POSTCODE	-	-	-	-
8	Exact baby's DOB + first four characters of POSTCODE	-	-	-	-
9	Month and year of Baby's DOB + exact POSTCODE	-	-	-	-

1 2005 - there were 51,645 (~8.36%) registration/NN4B records that did not match.

2 2006 - there were 55,852 (~8.72%) registration/NN4B records that did not match.

Source: HES

Linkage for babies' records consisted of five stages of an algorithm which included a mix of baby's NHS number, date of birth and postcode. The records were linked using all the stages (Table A2).

Table A2 **Number and percentage of registration/NN4B records that were linked to HES baby records by algorithm, 2005¹ and 2006²**

England

Stage	Variables used	Records linked			
		2005		2006	
		Number	Percentage	Number	Percentage
1	Exact SEX + Partial DOB ³ + Exact NHSNO	515,434	83.5	525,897	82.1
2	Exact SEX + Exact DOB + Exact POSTCODE	4,597	0.7	10,213	1.6
3	Exact DOB + Exact POSTCODE	176	<0.1	377	0.1
4	Exact DOB + first 4 characters of POSTCODE	319	0.1	1,838	0.3
5	Exact POSTCODE + month and year of DOB	1	<0.1	-	-

1 2005 - there were 97086 (15.7%) registration/NN4B records that did not match.

2 2006 - there were 101946 (15.9%) registration/NN4B records that did not match.

3 DOB = Baby's Date of Birth.

Source: HES

In some instances, after the first appropriate stage of the algorithm was applied leading to a HES record being linked to a registration and NN4B linked record, using further stages of the algorithm, the same HES record was linked again to a different registration and NN4B linked record. These were identified using the HESID. Records that were uniquely linked using either stages 1 or 2 of the algorithm were kept for analysis and the rest were examined manually to see whether data items such as date of birth of mother, date of birth of baby and postcode were consistent on HES and registration and NN4B linked file. Linkage to HES was deleted for records where the data items were inconsistent.

In addition, the data sent by Northgate Solutions contained multiple records for each HESID. These were identified using the variable Epikey. Multiple epikeys were present for both the mother and baby records. The record with the highest epikey was kept as this contained information for most of the variables required for analysis and all others were deleted.

Northgate Solutions had used the HES index to link the Maternity HES records to registration and NN4B linked data. Registration and NN4B linked records for 2005 and 2006 were linked using the most recent HES index and this might have contributed to some discrepancies in the linkage rate.