

Dattani, N., Datta-Nemdharry, P. & Macfarlane, A. J. (2012). Linking maternity data for England 2007: methods and data quality. *Health Statistics Quarterly*, 53, 4 - 21.



**CITY UNIVERSITY
LONDON**

[City Research Online](#)

Original citation: Dattani, N., Datta-Nemdharry, P. & Macfarlane, A. J. (2012). Linking maternity data for England 2007: methods and data quality. *Health Statistics Quarterly*, 53, 4 - 21.

Permanent City Research Online URL: <http://openaccess.city.ac.uk/1061/>

Copyright & reuse

City University London has developed City Research Online so that its users may access the research outputs of City University London's staff. Copyright © and Moral Rights for this paper are retained by the individual author(s) and/ or other copyright holders. Users may download and/ or print one copy of any article(s) in City Research Online to facilitate their private study or for non-commercial research. Users may not engage in further distribution of the material or use it for any profit-making activities or any commercial gain. All material in City Research Online is checked for eligibility for copyright before being made available in the live archive. URLs from City Research Online may be freely distributed and linked to from other web pages.

Versions of research

The version in City Research Online may differ from the final published version. Users are advised to check the Permanent City Research Online URL above for the status of the paper.

Enquiries

If you have any enquiries about any aspect of City Research Online, or if you wish to make contact with the author(s) of this paper, please email the team at publications@city.ac.uk.

Linking maternity data for England 2007: methods and data quality

Nirupa Dattani, Preeti Datta-Nemdharry and Alison Macfarlane *City University London*
(*School of Health Sciences*)

Abstract

Background

Maternity Hospital Episode Statistics (HES) data for 2007 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 extended the time period 2005–06, for which data had previously been linked and reported.

Methods

Birth registration and NN4B records were linked to Maternity HES delivery records and also to Maternity HES baby records using the NHS Number when available. Other direct identifiers were used if the 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 the quality of Maternity HES.

Results

Overall, 93 per cent of Maternity HES delivery records could be linked to the birth registration/NHS Numbers for Babies records and 80 per cent of Maternity HES baby records were linked to these.

Two per cent of Maternity HES records had the mother's NHS number missing compared with 22 per cent in the NN4B dataset. This did not reflect the extent to which other Maternity HES data items were missing or inconsistent between the two data sets.

Nearly a third of all linked Maternity HES records for singleton babies had one or more of the following data items missing: birth weight, 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 birth weight, 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, 75 per cent of the linked records had the same ethnic group recorded for the mother and her baby.

Conclusions

The linkage rate for 2007 was slightly higher than for the two previous years, but data were more incomplete. To gain maximum benefit from this linkage, improvements are urgently needed in the quality and completeness of the data contained in Maternity HES.

Contents

Abstract.....	1
Introduction.....	4
Methods.....	5
Results.....	6
Discussion.....	12
Conclusion.....	14
Acknowledgement.....	14
References.....	16
Appendix A.....	17

List of Tables

Table 1	Number of linked records for singleton births with missing data items in common data fields, 2007.....	8
Table 2	Comparison of plurality between birth registration and Maternity HES, 2007.....	8
Table 3	Comparison of live/ still birth status for singletons between birth registration and Maternity HES, 2007.....	9
Table 4	Comparison of sex for singletons in the linked records using birth registration and Maternity HES, 2007.....	9
Table 5	Comparison of birth weight distribution in the linked records using birth registration and maternity HES for singletons, 2007.....	11
Table A1	Comparison of gestational age in linked records using NHS numbers for babies and maternity HES for all births, 2007.....	17
Table A2	Comparison of baby's ethnicity from NHS Numbers for Babies with mother's ethnicity from HES, 2007.....	18

Introduction

The data recorded at birth registration are mainly socio-demographic: such as names, address of the mother's and father's usual place of residence, place of birth, occupations of the parents and dates of birth of the mother and baby (Office for National Statistics publication, DH3). As a result some key items needed for demographic and clinical purposes are not available at a national level. The opportunity to obtain gestational age and ethnicity data nationally resulted from 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 as babies born preterm, before 37 completed weeks of gestation, are at particularly high 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 be obtained only 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 was set up in 2004 between City University London, the Office for National Statistics (ONS) and the then Welsh Assembly Government 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; Moser K and Hilder L, 2008).

Stage 2 of the project involved linkage of the linked dataset for the years 2005 and 2006, created in stage 1, to Maternity HES and 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 were linked to Maternity HES and corresponding Welsh records at a later date using the experience gained in linking the first two years' data. The article published earlier describes details of the method used for linkage to Maternity HES records (Dattani *et al.*, 2011). This article reports on quality and completeness of the 2007 linked data. The Welsh linkage for all three years, 2005–07, will be reported separately.

Linkage of data for further years and access to the 2005–07 linked data for other projects will involve seeking approval from the ethics and permission from the National Information Governance Board to access individual patient identifiable records and securing new funding.

Several data items are common to all three data 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 is enabling new analyses using these linked data. For example, it is now possible to analyse caesarean section rates by the father's socio-economic classification, compare time of birth with birth outcomes, and report on the outcome of birth by onset of labour, gestational age, time of day and day of the week. Now 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		+	
Birth weight	+	+	+
Gestational age (still birth)	+	+	+
Gestational age (live birth)		+	+
Sex of baby	+	+	+
Number of babies born	+	+	+
Live or still birth	+	+	+
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

Details of the source data: birth registration, NHS Numbers for Babies and Maternity Hospital Episode Statistics (HES) can be found in the earlier article in Health Statistics Quarterly 49 describing the linkage of data for 2005 and 2006 (Dattani *et al.*, 2011).

Record linkage

Record linkage was carried out by Northgate Solutions, which processes HES records under contract with the NHS Information Centre. The linkage algorithm previously compiled for 2005 and 2006 data was used, but the program was slightly amended to ensure that only one HES record was linked to each registration-NN4B linked record (Dattani *et al.*, 2011).

The linked data provided to ONS by Northgate Solutions consisted of two files. One contained previously linked registration, NN4B records linked data to the mother's record in HES which also included the baby 'tails'. The second file based on linkage of registration, NN4B, linked records to

baby records in HES. These 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

The review of the quality of Maternity HES was focussed on the completeness and consistency of the HES data, in relation to birth registration data where 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, 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) The completeness of the main data items in all three sources was measured by identifying the extent to which data were missing.

The linked data for the mother's file was split into singleton and multiple births, using the multiple birth status field from registration, to facilitate the assessment of data quality. In some instances the results are reported separately.

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

Results

Mother file

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

Northgate solutions returned 630,409 records that had linked to the registration and NN4B linked data. These included some multiple records for the same mother for each episode. Records with the most complete information were selected to ensure one to one linkage to the registration and NN4B linked dataset. This gave a file of 615,239 records.

In the registration and NN4B linked data file, there were 659,061 records for babies who were either born in England or resident in England. The resident in England category was used for births recorded as occurring at home in the registration and NN4B linked data.

Around 73 per cent of the linked registration and NN4B records were linked to Maternity HES records using the mother's NHS number and her partial date of birth. A further 20 per cent of the linked registration and NN4B records were matched to Maternity HES using the mother's postcode and full date of birth. Only 7 per cent of registration and NN4B linked records were not linked to HES. A total of 614,369 Maternity HES records were linked to the registration and NN4B linked records giving a linkage rate of 93.2 per cent.

Baby file

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

A total of 667,893 HES baby records were linked to registration and NN4B linked data by Northgate solutions. This included multiple HES birth records for the same baby linked to a registration and NN4B linked record. Again only records with the fullest information were kept and others were deleted. After deletion, 552,398 records remained.

In the 2007 registration and NN4B linked data there were 659,061 records for babies who were either born in England or resident in England. Of these, 541,677 registration and NN4B linked records were linked to HES baby records using the NHS number, partial date of birth and sex, and 7,010 were linked using the baby's date of birth, postcode and sex. Over 16 per cent of registration and NN4B linked records could not be linked to HES baby records. Overall 552,313 of the 659,061 records were linked, giving a linkage rate of 83.8 per cent.

Data Quality

For HES, the extent to which data were missing or discordant was assessed only in the mother's records as these included information on the baby and also because the linkage rate was far better than for the baby records. For multiple births, information was recorded only for 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, 22 per cent of linked registration and NN4B records did not have the mother's NHS number compared with 2 per cent in the Maternity HES records. In Maternity HES, birth weight and gestational age information was missing for 31 per cent and 47 per cent of singletons respectively. Information about live or still birth status and/or the baby's date of birth and sex was missing in nearly a third of the records (Table 1).

Table 1 Number of linked records¹ for singleton births with missing data items in common data fields, 2007

England

	NN4B		Birth registration		Maternity HES	
	Number	Percentage	Number	Percentage	Number	Percentage
NHS number of mother	131,202	22.0	NA	NA	11,546	1.9
Date of birth of mother	0	0.0	2,066	0.3	0	0.0
Ethnicity	51,975	8.7	NA	NA	80,283	13.5
Postcode	0	0.0	23	0.0	1,984	0.3
Birth weight	2,942	0.5	5,177	0.9	184,212	30.9
Gestational age	6,323	1.1	NA	NA	281,338	47.3
Status	0	0.0	0	0.0	194,811	32.7
Date of birth of baby	0	0.0	0	0.0	187,931	31.6
Sex of baby ²	763	0.1	0	0.0	196,545	33.0

1. Out of 595,371 singletons

2. Includes 763 cases with indeterminate sex in NN4B and in Maternity HES 368 cases with indeterminate sex, 66 cases coded to 4 and 5 cases coded to 5,7 and 8.

Source: HES, registration and NHS numbers for Babies

Discordances in the 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. Where data items were not recorded at birth registration, NN4B data were used.

Discordance in multiple birth status

There were 14,274 records identified as relating to multiple births in birth registration and Maternity HES. Multiple birth status was discordant between the two data sources in 3,205 records (Table 2).

Table 2 Comparison of plurality between birth registration and Maternity HES, 2007

England

Year	Maternity HES	Birth registration		Total
		Singleton	Multiple	
2007	Singletons	464,464	524	464,988
	Multiple	2,681	14,274	16,955
	Not stated	128,226	4,200	132,426
	Total	595,371	18,998	614,369

Source: HES and registration

Discordance in live or still birth status

For the records which had a stated live or still birth status in both data sources, one per cent of the records disagreed on birth status (Table 3). Around 33 per cent of linked Maternity HES records had no information on birth status.

Table 3 Comparison of live/ still birth status for singletons between birth registration and Maternity HES, 2007

England

		Birth registration			% of all records
		Live birth Number	Still birth Number	Total Number	
Maternity HES	Live	394,780	80	394,860	66.3
	Still birth: ante-partum	4,006	1,354	5,360	0.9
	Still birth: intra-partum	6	154	160	0.0
	Still birth: Indeterminate	12	168	180	0.0
	Not Known	193,777	1,034	194,811	32.7
	Total		592,581	2,790	595,371

Source: HES and registration

Discordance in baby's sex

The sex of the baby recorded on birth registration for singleton births was compared with Maternity HES. Where the baby's sex was recorded in both data sources, an agreement of 98 per cent was observed (Table 4). Sex was indeterminate in 763 in NN4B records and 368 cases in Maternity HES. In the latter, sex was coded to unspecified codes in 71 cases (as shown in the footnote in Table 4).

Table 4 Comparison of sex for singletons in the linked records using birth registration and Maternity HES, 2007

England

		Birth registration			% of total
		Male	Female	Total	
Maternity HES ¹	Male	201,224	1,806	203,030	34.1
	Female	4,209	191,587	195,796	32.9
	Not stated	101,062	95,483	196,545	33.0
	Total	306,495	288,876	595,371	100.0

1. Includes 368 cases with indeterminate sex, 66 cases coded to 4 and 5 cases coded to 5,7 and 8.

Source: HES and registration

Discordance in birth weight

Where birth weight was recorded, there was good concordance between Maternity HES and birth registration. In Maternity HES, birth weight was missing in a third of the records, however, compared to only 1 per cent in birth registration (Table 5).

Table 5 Comparison of birth weight distribution in the linked records using birth registration and maternity HES for singletons, 2007

England															
Maternity HES		Birth registration											Total	%	
Birth weight(g)	<500	500-999	1000-1499	1500-1999	2000-2499	2500-2999	3000-3499	3500-3999	4000-4499	4500-4999	5000-5499	5500 and over	Not stated	Total	%
<500	188	14	8	5	11	30	39	47	19	5	0	3	11	380	0.1
500-999	4	1,561	4	3	2	9	5	12	0	2	1	11	38	1,652	0.3
1000-1499	2	2	2,196	9	8	6	7	6	8	0	0	5	55	2,304	0.4
1500-1999	3	2	10	4,401	10	7	14	9	1	1	0	2	62	4,522	0.8
2000-2499	4	2	4	8	15,590	70	80	17	10	3	0	1	144	15,933	2.7
2500-2999	8	5	9	10	174	66,133	116	165	12	4	3	5	520	67,164	11.3
3000-3499	15	15	21	9	104	121	147,845	379	64	6	3	3	1,201	149,786	25.2
3500-3999	13	17	13	8	26	169	463	120,203	48	15	1	10	897	121,883	20.5
4000-4499	12	3	9	2	7	15	52	51	39,681	44	3	4	269	40,152	6.7
4500-4999	2	0	1	1	0	1	9	13	16	6,482	0	0	46	6,571	1.1
5000-5499	0	4	0	0	1	0	3	0	1	0	663	1	6	679	0.1
5500 and over	0	24	1	0	4	7	17	14	6	0	1	59	0	133	0.0
Not stated	178	815	1,113	2,125	7,141	30,084	66,371	53,549	17,744	2,819	299	46	1,928	184,212	30.9
Total	429	2,464	3,389	6,581	23,078	96,652	215,021	174,465	57,610	9,381	974	150	5,177	595,371	100.0
%	0.1	0.4	0.6	1.1	3.9	16.2	36.1	29.3	9.7	1.6	0.2	0.0	0.9	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 nearly 90 per cent of the records where it was recorded in both sources, gestational age was the same (see Table A1 in the Appendix). On the other hand, in Maternity HES, almost half of all records had gestational age missing. Gestational age differed by one week in around 6 per cent of the records and two weeks or more in about 9 per cent of the records. There was a wide variation in gestational age between the two data sources in the 'tails' for babies born before 22 weeks and over 42 weeks, but only 4 per cent of births occurred in these extremes of the gestational age distribution. The difference was 23 per cent for those born before 22 weeks. At 42 weeks, gestational age differed in about a fifth of all records. For records of births at 43 weeks or over, gestational age was missing in 43 per cent of maternity HES records.

Discordance in ethnicity

The baby's ethnicity recorded in the NN4B record and the mother's ethnicity recorded in Maternity HES were compared (see Table A2 in the Appendix). There was agreement in three-quarters of the records which had a stated ethnic category. Among all the linked records, 13 per cent of records had no ethnicity recorded in Maternity HES and in 9 per cent of records ethnic group was not stated in the NN4B data.

Discussion

Three-quarters of the registration and NN4B records were linked to the HES mothers' records using the NHS number and partial date of birth. This was not surprising as the mother's NHS number was missing from nearly a quarter of the registration and NN4B linked records, and also from a very small proportion of Maternity HES records. A further fifth of the registration and NN4B linked records were linked using the date of birth or month and year of birth, and the postcode. There were concerns about using postcodes 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 were 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. This could have been improved further if there had been a shorter delay before linkage was carried out as HESID would have been less likely to have changed. Alternatively HESID at birth could be retained as a separate field for linkage. There are however about 20 Trusts that fail to submit any maternity data to HES because they have a stand-alone maternity system that is not linked to the Patient Administration System. Hence it would be impossible to obtain a much higher linkage rate until all Trusts in England submit data to HES.

The linkage rate for registration and NN4B linked records to HES baby records was slightly lower than the linkage rate for the mothers' records. This was not surprising, as a large proportion of baby 'tails' are known to be missing in Maternity HES (HES website 2010).

HES mother records include information about the baby. As the linkage rate for registration and NN4B linked data to HES mother records was higher than for the baby records, the quality of information in HES was assessed using the mothers' records. There were however issues with multiple births in the HES mothers' record, as already found in the 2005/06 data. Multiple birth status was also unknown in a fifth of the records. Further work is needed to assess the quality of

data on multiple births for all three years of linked data before they could be used for any analyses.

Discrepancy in the recording of live/stillbirth status for singleton babies was found in 1 per cent of the linked records. This shows a deterioration compared with the data for the two previous years where it was 5 in 100,000 records in 2005 and 2 in 1,000 records in 2006. A third of the HES records for 2007 did not have any information on birth status, which is consistent with the 2005 and 2006 data.

Birth weight was missing in a quarter of all linked Maternity HES records for singleton babies compared with only 0.2 per cent at birth registration. There was however, good concordance between the two data sources where birth weight was stated, as the majority of the records were in the same 500g birth weight group. Missing birth weights are investigated by ONS by going back to registrars and also to child health departments. Therefore the quality of birth weight information on birth registration is 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 the UK. 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 their menstrual history recorded (Confidential Enquiry into Stillbirths and Deaths in Infancy report, 2001).

Gestational age distributions have shown to differ according to 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 lower, but recorded gestational age differed by one week in only 7 per cent of the linked records. Nearly half of the linked HES records had no information about gestational age, compared with only 1 per cent in the NN4B data. Sub-national analysis of the NN4B data for 2005–08 showed that majority of the Trusts had none or very few records with gestational age missing (Office for National Statistics publication, Quality of ethnicity and gestational age data for 2005–08). Where gestational age was stated in maternity HES, it was in good agreement with NN4B in majority of the records.

A past study using maternity HES data for 1990–91 showed that only 52 per cent of the deliveries were recorded on HES compared with the number of registered births and, within regions, the level of completeness varied from district to district (Middle C, Macfarlane A, 1995). There has been a vast improvement in the number of maternities recorded on HES since that time but the level of completeness still varies between NHS Trusts (NHS Information Centre, Maternity HES Statistics bulletin 2007–08).

The NN4B system records information about the ethnic category of the baby as defined by the mother, using the 2001 Census categories (Moser K, Stanfield KM, *et al.*, 2008). On the Maternity HES record, the mother's ethnicity is self-reported using the 2001 Census categories. It is unclear however, whether the mother was involved in defining the ethnic category in either of these data

sources or whether a health professional decided what to record without asking the mother. In practice it is likely to be a mixture of both. Although the ethnic group of the baby is requested in NN4B, it is not possible to know whose ethnic group 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. Individual responses, whether self-reported or not, may vary according to circumstances and over time.

Despite these limitations, in three-quarters of the linked records the mother's ethnicity recorded was the same as that recorded for her baby. In 3 per cent of records, the mother's ethnicity was categorised as 'White British' and baby's ethnicity was categorised as 'White other' or vice versa. This suggests that the 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. Although recording of ethnicity is better on NN4B than in maternity HES, the level of completeness varies by Trusts ranging from zero to 98 per cent (Office for National Statistics publication, Quality of ethnicity and gestational age data for 2005–08).

Conclusion

This study shows that it is possible to link the majority of the Maternity HES records routinely to registration and NN4B linked records, but linkage would be considerably more valuable if there were further improvements in the quality and completeness of Maternity HES. Information about method of delivery and complications in pregnancy can only be obtained at a national level from Maternity HES, so linkage would be needed 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. On the other hand, where data have been recorded they are in good concordance with birth registration or NN4B but there are 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.

References

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.

Dattani N, Datta-Nemdharry P, and Macfarlane A. (2011) Linking maternity data for England, 2005–06: methods and data quality. *Health Statistics Quarterly* 49. Available on the ONS website at: www.ons.gov.uk/ons/rel/hsg/health-statistics-quarterly/spring-2011/index.html

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.

NHS Information Centre, *Maternity data, 2007–08*. Available at:

www.hesonline.nhs.uk/Ease/servlet/ContentServer?siteID=1937&categoryID=1060

Office for National Statistics, *Mortality Statistics: Childhood, infant and perinatal, England and Wales, 2007*. Series DH3 No. 40. Available on the ONS website at:

www.ons.gov.uk/ons/rel/vsob1/mortality-statistics--childhood--infant-and-perinatal--england-and-wales--series-dh3-/no--40--2007/index.html

Office for National Statistics, *Quality of ethnicity and gestation data subnationally for births and infant deaths in England and Wales, 2005–08*. Available on the ONS website at:

www.ons.gov.uk/ons/publications/re-reference-tables.html?edition=tcm%3A77-226528

Appendix A

Table A1 - Comparison of gestational age in linked records using NHS numbers for babies and maternity HES for all births, 2007

Maternity HES	NHS Numbers for Babies																							Total			
Gestational age (weeks)	<22	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43 44 and over	Not stated	Total		
<22	75	1	1	12	6	7	2	10	14	10	21	0	1	1	1	2	2	2	3	10	4	0	0	0	2	187	
22	0	50	2	1	0	0	0	0	0	0	0	21	0	0	1	0	0	2	0	1	2	0	0	0	0	80	
23	1	1	108	2	0	0	0	0	1	0	0	0	14	55	1	0	1	1	2	1	1	0	0	0	2	191	
24	0	0	5	280	6	1	3	0	0	0	1	0	1	64	3	0	1	1	1	1	1	0	0	0	0	369	
25	0	0	5	9	309	10	2	0	0	2	0	0	0	1	96	216	5	2	8	3	0	0	0	0	1	669	
26	0	0	0	3	2	380	6	5	4	2	0	1	0	0	1	4	556	8	5	13	2	0	0	0	2	994	
27	0	0	0	0	0	8	382	8	3	0	0	1	1	2	0	3	1	15	1,036	25	20	2	0	0	5	1,512	
28	0	0	0	1	2	2	6	511	8	1	0	0	0	4	3	3	13	22	32	1,224	1,120	10	0	0	7	2,969	
29	2	0	0	1	0	0	1	3	551	14	0	2	1	1	0	2	3	12	28	13	15	304	2	0	7	962	
30	2	0	0	0	0	2	1	2	11	619	26	7	1	2	3	5	0	3	9	8	2	2	7	0	14	726	
31	0	1	0	0	0	1	2	2	5	22	825	27	9	1	3	5	1	3	4	4	4	1	0	0	15	935	
32	0	0	0	0	0	2	2	2	1	6	18	1,261	25	8	6	8	7	6	8	11	8	1	0	0	18	1,398	
33	0	0	0	0	0	0	0	1	3	5	6	41	1,702	52	20	3	10	14	8	13	6	1	0	0	22	1,907	
34	0	0	0	1	2	1	1	0	0	1	1	7	54	2,778	79	23	21	30	26	16	10	0	1	1	43	3,096	
35	0	0	0	0	0	0	0	0	0	1	0	0	6	14	69	4,101	138	47	44	49	40	19	0	2	1	64	4,595
36	1	0	1	0	0	2	1	2	1	1	3	3	7	18	144	7,817	319	100	86	90	49	6	8	1	120	8,780	
37	1	0	0	2	0	0	6	1	2	1	1	5	3	11	32	230	17,128	762	207	123	69	33	24	2	261	18,904	
38	5	0	0	1	1	2	1	3	1	2	2	3	9	21	29	87	563	40,119	1,612	348	121	42	28	10	453	43,463	
39	1	0	0	0	3	1	2	1	5	7	1	2	4	18	31	40	169	1,358	64,831	2,243	322	78	55	25	749	69,946	
40	8	1	1	0	0	1	5	3	1	5	3	3	5	9	9	56	132	497	2,320	79,875	3,382	389	76	51	814	87,646	
41	1	0	0	1	1	0	2	3	0	2	1	2	3	7	8	30	71	130	398	2,431	55,647	1,073	80	39	560	60,490	
42	1	0	0	0	0	0	0	0	0	0	0	3	2	1	2	11	24	70	166	495	1,309	9,829	37	17	150	12,117	
43	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	3	10	35	94	136	262	80	498	0	2	1,121	
44 and over	0	0	0	0	0	0	0	0	0	0	0	0	0	6	2	4	20	50	115	226	260	49	8	236	4	980	
Not stated	132	87	139	275	343	368	415	569	605	728	938	1,357	1,797	2,869	4,292	7,887	16,768	38,468	65,005	76,260	56,076	10,803	558	368	3,225	290,332	
Total	230	141	262	589	675	788	840	1,126	1,217	1,428	1,846	2,753	3,652	5,934	8,833	16,457	35,530	82,305	136,050	163,607	118,725	22,706	1,384	751	6,540	614,369	

Source: HES and NHS Numbers for Babies

Table A2 - Comparison of baby's ethnicity from NHS Numbers for Babies with mother's ethnicity from HES, 2007

NHS Numbers for Babies	Maternity HES																	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	322,195	1,319	9,485	296	99	309	576	248	256	48	202	175	226	120	124	1,848	50,392	387,918
White Irish	746	978	81	3	0	4	1	4	1	2	0	1	2	0	0	29	204	2,056
Any other White	5,278	207	24,964	42	66	63	624	32	36	8	222	30	89	44	41	2,402	4,477	38,625
White and Black Caribbean	2,695	20	178	785	34	11	171	4	2	2	11	484	87	141	2	120	633	5,380
White and Black African	1,259	12	256	42	416	6	58	14	3	1	12	45	804	143	0	137	492	3,700
White and Asian	2,060	33	309	9	5	371	96	482	150	50	495	11	11	10	157	349	782	5,380
Any other mixed	2,646	45	1,072	320	101	111	840	222	123	17	430	213	169	169	271	865	1,276	8,890
Indian	299	2	50	11	6	58	41	12,020	861	284	1,152	15	39	19	6	307	1,828	16,998
Pakistani	324	2	36	2	3	83	38	501	20,185	289	936	6	21	26	1	308	2,251	25,012
Bangladeshi	92	1	10	3	3	11	6	178	469	7,105	254	2	10	7	2	90	478	8,721
Any other Asian	173	0	146	7	11	90	78	770	486	140	4,283	17	73	52	196	1,197	1,040	8,759
Caribbean	409	3	44	341	37	2	62	10	5	3	8	4,295	299	575	2	189	674	6,958
African	280	5	95	58	562	12	115	31	59	12	148	389	14,610	994	3	711	2,435	20,519
Any other Black	168	4	62	68	85	9	67	18	19	3	64	379	791	2,818	2	207	498	5,262
Chinese	44	0	27	0	0	12	39	4	3	0	118	0	3	1	2,040	155	407	2,853
Any other	1,466	23	1,701	91	101	93	292	188	208	30	1,180	103	396	140	216	4,978	1,885	13,091
Not stated	31,112	172	2,239	210	97	113	344	2,240	996	586	666	368	1,252	224	221	1,241	12,166	54,247
Total	371,246	2,826	40,755	2,288	1,626	1,358	3,448	16,966	23,862	8,580	10,181	6,533	18,882	5,483	3,284	15,133	81,918	614,369

Source: HES and NHS Numbers for Babies