



City Research Online

City, University of London Institutional Repository

Citation: Datta-Nemdharry, P., Dattani, N. & Macfarlane, A. J. (2012). Linking maternity data for Wales, 2005-07: methods and data quality. *Health Statistics Quarterly*, 54(54), pp. 1-24.

This is the accepted version of the paper.

This version of the publication may differ from the final published version.

Permanent repository link: <https://openaccess.city.ac.uk/id/eprint/6944/>

Link to published version:

Copyright: City Research Online aims to make research outputs of City, University of London available to a wider audience. Copyright and Moral Rights remain with the author(s) and/or copyright holders. URLs from City Research Online may be freely distributed and linked to.

Reuse: Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

Linking Maternity Data for Wales, 2005–07: Methods and Data Quality

Preeti Datta-Nemdharry, Nirupa Dattani and Alison Macfarlane, City University London (School of Health Sciences) and Gwyneth Thomas, Welsh Government (Health Statistics and Analysis Unit)

The Office for National Statistics (ONS) welcomes articles which inform readers on new developments in methodology affecting official statistics on health, and promote the wider use and understanding of research based on ONS data and other official statistics. All articles are peer reviewed in accordance with international standards for scientific journals. Publication does not imply endorsement for National Statistics purposes by ONS.

Abstract

Background

This work formed part of a project to link data recorded routinely at birth in England and Wales to bring together socio-demographic data and data about care at birth. Birth registration data for England and Wales had already been linked to the National Health Service (NHS) Numbers for Babies' data (NN4B) recorded when an NHS number is issued to a new baby. The data for babies born between 2005 and 2007 to mothers resident in England were then further linked to their records in the Maternity Hospital Episode Statistics (HES). This paper describes the linkage of linked birth registration and NN4B records for babies of mothers resident in Wales for 2005 to 2007 to the Patient Episode Database for Wales (PEDW) and the National Community Child Health Database (NCCHD) records.

Methods

Birth registration and NN4B records were first linked to NCCHD records, which contains data about the children born. This dataset was further linked to PEDW to obtain maternity records relating to their delivery. The linkage was carried out using pre-defined linkage algorithms.

The quality of the Welsh data was assessed in terms of completeness of data and concordance of common data items in relation to birth registration wherever possible. NN4B data were used for data items not collected at birth registration, such as gestational age for live births.

Results

Around 92 per cent of registration/NN4B records for the three years could be linked to NCCHD and PEDW records. Different data fields were provided from each of the two Welsh data sources, with the common and key data items, such as birth weight and gestational age, coming from NCCHD.

Overall the percentage of missing data in NCCHD was minimal, with the exception of ethnicity which was missing from 13 per cent of records in NCCHD in 2005 but from fewer records in subsequent years.

For births in 2005 and 2006, over 30 per cent of NN4B records linked to registration did not have the mother's NHS number compared with less than 1.1 per cent of NCCHD records.

There was excellent agreement between the data items in linked birth registration and NN4B files and the data from NCCHD with over 95 per cent concordance in common data items. Around 99 per cent of the linked records had the same ethnic group which is not surprising as records on the child health system databases and NN4B are derived from a common data source.

Conclusion

The linkage rate for maternity data in Wales was similar to that obtained in linking registration/NN4B linked data to the Maternity HES records for England but the data were of higher quality and were more complete. Therefore, NCCHD linked to PEDW could be used to analyse birth outcomes for Wales without the need to link to birth registration and NN4B data. Nevertheless data items such as mother's country of birth and socio-economic status are recorded only at birth registration so linkage to the birth registration/NN4B dataset can generate a much fuller set of data items and enable analyses of birth outcomes by factors such as ethnicity, socio-economic status and parents' country of birth.

This article can be cited as: Dattani N, Datta-Nemdharry P, Macfarlane A, Thomas G., Linking maternity data for Wales 2005-07: methods and data quality., Health Stat Q. 2012 Summer;(54):1-24.

Contents

Abstract.....	1
Introduction.....	5
Methods.....	6
Results.....	9
Discussion.....	20
Conclusion.....	21
Acknowledgement.....	21
References.....	22
Appendix A.....	24

List of Tables

Table 1a	Number of linked records with missing data items in common data fields, 2005.....	10
Table 1b	Number of linked records with missing data items in common data fields, 2006.....	10
Table 1c	Number of linked records with missing data items in common data fields, 2007.....	11
Table 2	Comparison of plurality for linked births between registration and NCCHD, 2005, 2006 and 2007.....	12
Table 3	Comparison of birth status for linked births between registration and NCCHD, 2005, 2006 and 2007.....	13
Table 4	Comparison of sex for linked live births between registration and NCCHD, 2005, 2006, 2007.....	14

Table 5a	Comparison of birth weight distribution for linked births between registration and NCCHD, 2005.....	15
Table 6a	Comparison of gestational age for linked births between NN4B and NCCHD, ...	17
Table 7a	Comparison of baby's ethnicity from NN4B and NCCHD for linked births, 2005.....	19
Table A1	Number and percentage of registration/NN4B/NCCHD records that were linked to PEDW records by algorithm, 2005–07.....	24

Introduction

This work formed part of a project to link data recorded routinely at birth in England and Wales, to bring together socio-demographic data and data about care at birth. At birth registration vital information such as names of parents and babies; address of residence; place of birth; birth weight; date of birth; occupations of parents; marital status of parents; and parents' countries of birth are recorded, but other key data items needed for clinical and demographic purposes, for example gestational age, method of delivery and ethnicity, are not. Gestational age is an important birth outcome and babies born before 37 weeks of gestation are more likely to die in the early years of life or experience immediate and long-term morbidity than those born later (Brocklehurst, 1999; Confidential Enquiry into Maternal and Child Health, 2004; European Perinatal Health report, 2008). The 'NHS Numbers for Babies' service (NN4B), introduced in 2002, retains a small set of data recorded about birth, including gestational age and baby's ethnicity.

The Maternity Hospital Episode Statistics (HES) dataset for births that have occurred in England, and Patient Episode Database for Wales (PEDW) and National Community Child Health Database (NCCHD) for births that have occurred in Wales, provide clinical information on maternity care at delivery at a national level. Therefore, in 2004, City University London; the Office for National Statistics (ONS); the Welsh Government; and other relevant organisations in England and Wales collaborated to link these datasets for all births that occurred in England and Wales from 2005 to 2007. The first phase of the project involved linkage of birth registration data with the NN4B dataset (Hilder *et al*, 2007). In addition, the quality and completeness of the NN4B data were assessed (Moser and Hilder, 2008). Subsequently, an assessment of the quality and completeness of ethnicity and gestational age data from NN4B for the years 2005 to 2008 at a sub-national level was published in an ONS statistical bulletin (ONS, 2011).

To extend the range of data available for England, the linked birth registration and NN4B dataset created in the first phase was further linked to Maternity HES for the years 2005 and 2006 (Dattani *et al*, 2011). At that time the 2007 linked birth registration and NN4B dataset was not yet available but this has now also been linked to the 2007 Maternity HES dataset (Dattani *et al*, 2012).

This article describes linkage of the linked birth registration and NN4B dataset to PEDW and NCCHD for the years 2005 to 2007, to extend the range of data available for Wales. It also assesses the quality and completeness of the linked data.

Box 1 shows the data items that are common among the four data sources and those that are unique to each data source. In addition to patient identifiable information used for linkage, some data items are recorded on both NCCHD and PEDW. The main purpose of this linkage and that done in the previous phase is to enable new and wider ranging analyses, drawing on all the data sources.

Box 1 Selected data items from birth registration, NN4B and PEDW and NCCHD				
	Data sources			
Data items	Birth registration	NN4B	NCCHD	PEDW*
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				+

*Some data items are recorded on the mother's record and some on the baby's record.

Methods

Source data

Birth registration

In England and Wales, there is a legal requirement to register all live births within 42 days of birth. The process of registration, the data items collected, the legal basis and the definition of a live birth are described in detail elsewhere (ONS, 2010). The local registrar of births, marriages and death records the information from one or both parents. Some information is passed on from the local child health department to the registrar to verify the birth. Since 1975 this has included the baby's birth weight (National Health Service Act 1977) and since 2002 the NHS number, but not the gestational age.

When a still birth is registered, the informant also gives the registrar a medical certificate of still birth completed by the midwife or doctor who attended the birth. The certificate includes the cause of death and an assessment of gestational age at birth and birth weight. The registrars do not retain the baby's NHS number for still births.

NHS Numbers for Babies (NN4B) service

The NHS Number for Babies (NN4B) dataset is generated when the National Health Service (NHS) number, a national unique patient identifier, is issued. Since 2002, an electronic notification of each birth has been sent to the Central Issue System so that the NHS number for the baby can be generated and a small set of data is recorded about the birth. This includes the gestational age at birth and the baby's ethnicity. The records are deleted from the system after six months. When the NHS number is issued, a copy of the information is sent to the local child health system where the birth occurred. A limited dataset, including the NHS number, is sent to the NHS Central Register. In 2005, the Office for National Statistics (ONS) started to receive a subset of the variables from this dataset for linkage to birth registration data. A pilot linkage of NN4B with birth registration data for 2005 was carried out (Hilder *et al*, 2007), after which these two data sources have been routinely linked.

Patient Episode Dataset Wales (PEDW)

PEDW is a database of individual hospital patient records. It includes all inpatient and day case activity in NHS Wales and data on Welsh residents treated in English NHS Trusts. In 1997, the mandated inpatient and day case dataset was changed to the Admitted Patient Care (APC) record used in Maternity HES. The decision to adopt the APC was to align the Welsh inpatient and day case dataset with England to allow benchmarking. The APC contains demographic, clinical and administrative items, such as the age and sex of the patient; diagnostic and operative procedures undertaken during the episode of care in hospital.

The APC extracts received from the Welsh NHS organisations and from English NHS organisations that treat Welsh residents are used to update PEDW.

Individual records are submitted to PEDW on the basis of a patient's consultant episode, which is the care an admitted patient receives in the continuous care of one consultant within one Local Health Board. If the patient is transferred to the care of another consultant, either in the same or another specialty or if they are transferred to another Local Health Board for continuing in-patient care, another consultant episode will start and another PEDW record will be generated (NHS Wales Informatics Service, 2009).

Patients whose episodes are captured by the PEDW database are classified as in-patients, day cases, maternity patients and regular attendees. Maternity patients comprise of a pregnant or recently pregnant woman being admitted to a maternity ward (including delivery facilities). Every hospital birth in Wales should have a PEDW record. The record consists of general information, diagnosis and procedure information for the mother. A maternity tail, in other words the delivery record, is attached to each general record for each baby born and contains fields with data relating to the relevant mother and babies. The completeness and quality of data in the maternity tail of the record are known to be poor (Welsh Government, 2012).

National Community Child Health Database (NCCHD)

The National Community Child Health Database (NCCHD) is Wales' first national community child health system and contains anonymised records for all children born after 1987 who were resident or treated in Wales. It has been created by bringing together selected information from locally managed community child health databases. Since 2002, community child health database records have been initiated from NN4B records but may subsequently be amended locally. The database is used to produce maternal and child health statistics for Wales as well as supporting the administration of child immunisation and health surveillance programmes.

Information about gestational age; onset of labour; the number of babies born; each baby's birth weight; sex; birth status (live or still birth); mode of delivery; ethnicity; birth location; breastfeeding at birth; and breastfeeding at 6/8 weeks were obtained from NCCHD. Clinical and hospital related data items were obtained from the PEDW dataset.

Record linkage

A subset of each data record, containing the key identifying variables shown in Appendix A1 and used in the linkage algorithm, was sent by the Office for National Statistics (ONS) to the NHS Wales Informatics Service (NWIS), formerly Health Solutions Wales, for linkage to NCCHD and PEDW.

The NCCHD contains records for individual children whereas maternity records in PEDW relate primarily to mothers. As the birth registration/NN4B dataset contains individual baby records, these were first linked to NCCHD. The mother's NHS number in NCCHD was then used in further linkage to mothers' records in PEDW. Full details of the linkage of registration and NN4B linked data to NCCHD and PEDW data are shown in Appendix A.

The linked data are held at ONS and researchers from City University London accessed the data via the secure data environment of the Virtual Microdata Laboratory (VML). ONS released the outputs of analyses in the form of disclosure controlled tables.

Data quality

The key data items that were evaluated as part of this paper came from NCCHD. The quality of NCCHD was assessed in terms of completeness of data and concordance of common data items in relation to birth registration or NN4B data. Birth registration was the main dataset against which consistency of NCCHD was compared since the information collected at registration is subject to quality checks (ONS, 2010). NN4B data were used to validate NCCHD where information was not available from birth registration.

In this paper some results are reported, mainly for 2005. The full set of results for 2005 to 2007 are available at: www.ons.gov.uk/ons/publications/re-reference-tables.html?edition=tcn%3A77-262405

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

Results

Linkage rates

In the first stage of the linkage, 32,626 NCCHD records were linked to birth registration/NN4B linked data for 2005. Of these, 30,620 records were matched with maternity data in PEDW. Therefore a total of 30,620 linked records were provided by NWIS. For 2006 and 2007, 31,816 and 32,284 records were linked respectively. These included multiple records for the same mother for each episode. Records with the fullest information were selected to ensure one-to-one linkage. There were some records where the episodes start and end dates were later than the date of birth and so these were deleted. This gave a file of 30,256; 31,498; and 32,141 records for 2005, 2006 and 2007 respectively.

The linked registration and NN4B dataset contained 31,758 babies, born in 2005, who were either born in Wales or resident in Wales. There were 32,798 and 33,592 babies born in 2006 and 2007 respectively. The data sent by NWIS included mothers resident in Wales with babies born in either England or Wales and also mothers resident in England with babies born in Wales. The latter were deleted as these were captured in the Maternity HES dataset (Dattani *et al*, 2011). Therefore, for 2005, 29,082 records in NCCHD and PEDW were linked to birth registration/NN4B data giving a linkage rate of 91.6 per cent. The linkage rates for 2006 and 2007 were 92.3 per cent and 92.0 per cent respectively.

The first phase of the linkage involved linking birth registration/NN4B linked records to NCCHD, using the baby's NHS number which was available on all records in both the datasets. For 2005, 97.5 per cent of the records in the linked registration/NN4B/NCCHD dataset were linked to PEDW data using the mother's NHS number in stage 1 of the algorithm. A further 719 of the linked registration/NN4B/NCCHD records were matched to records in PEDW using combinations of the mother's date of birth and post code. Around 8.4 per cent of linked registration and NN4B records were not linked to NCCHD and PEDW. Similar percentages were observed for 2006 and 2007 (see Appendix A).

Data quality

Missing data

For 2005, 37.8 per cent of NN4B records linked to registration did not have the mother's NHS number compared with 1.1 per cent of NCCHD records. Ethnicity was missing in 13 per cent of the linked records in NCCHD. This decreased to 6.8 per cent in 2006 and 1.3 per cent in 2007. Overall the percentage of missing data in other fields in linked NCCHD records was less than 1 per cent, as shown in Tables 1a, 1b and 1c.

Table 1a Number of linked records¹ with missing data items in common data fields, 2005

	NHS Numbers for babies		Registration		NCCHD	
	Number	Percentage	Number	Percentage	Number	Percentage
NHS number of mother	10,990	37.8	NA		319	1.1
Date of birth of mother	1	<0.1	205	0.7	3	<0.1
Ethnicity	316	1.1	NA		3,914	13.5
Postcode	825	2.8	2	<0.1	2	<0.1
Birth weight	228	0.8	5	<0.1	11	<0.1
Gestational age	29	0.1	NA		134	0.5
Status	0	0.0	0	0.0	0	0.0
Date of birth of baby	0	0.0	0	0.0	0	0.0
Sex of baby	26 ²	0.1	0	0.0	20 ³	<0.1

1. Number of linked records = 29,082.

2. Indeterminate sex.

3. Out of 28,950 live births.

Source: Registration, NHS Numbers for Babies and NCCHD

Table 1b Number of linked records¹ with missing data items in common data fields, 2006

	NHS Numbers for babies		Registration		NCCHD	
	Number	Percentage	Number	Percentage	Number	Percentage
NHS number of mother	9,135	30.2	NA		225	0.7
Date of birth of mother	3	<0.1	487	1.6	2	<0.1
Ethnicity	147	0.5	NA		2,064	6.8
Postcode	801	2.6	1	<0.1	1	<0.1
Birth weight	237	0.8	12	<0.1	22	<0.1
Gestational age	34	0.1	NA		160	0.5
Status	0	0.0	0	0.0	0	0.0
Date of birth of baby	0	0.0	0	0.0	0	0.0
Sex of baby	17 ²	<0.1	0	0.0	12 ³	<0.1

1. Number of linked records = 30,290.

2. Indeterminate sex.

3. Out of 30,154 live births.

Source: Registration, NHS Numbers for Babies and NCCHD

Table 1c Number of linked records¹ with missing data items in common data fields, 2007

	NHS Numbers for babies		Registration		NCCHD	
	Number	Percentage	Number	Percentage	Number	Percentage
NHS number of mother	²		NA		138	0.4
Date of birth of mother	1	<0.1	12	<0.1	1	<0.1
Ethnicity	180	0.6	NA		401	1.3
Postcode	²		²		0	0.0
Birth weight	157	0.5	218	0.7	32	0.1
Gestational age	40	0.1	NA		118	0.4
Status	0	0.0	0	0.0	0	0.0
Date of birth of baby	0	0.0	0	0.0	0	0.0
Sex of baby	24 ³	0.1	0	0.0	36 ⁴	0.1

1. Number of linked records = 30,895.

2. Not provided by ONS.

3. Indeterminate sex.

4. Out of 30,746 live births.

Source: Registration, NHS Numbers for Babies and NCCHD

Discordance

Discordance in common individual data items

The baby's date of birth was discordant between registration and NCCHD in 0.13 per cent of the linked records for babies born in 2005. This was similar for the births in 2006 and 2007. The postcodes of mother's usual place of residence were 100 per cent concordant between the data sources.

Discordance in multiple birth status

In the linked birth registration and NCCHD data for 2005, 786 babies were identified as being born as multiple births. This figure was 820 in 2006 and 898 in 2007. Generally, there was good agreement between birth registration and NCCHD about their multiple birth status. For all three years, multiplicity was discordant in less than 0.2 per cent of all the records (Table 2).

Table 2 Comparison of plurality for linked births between registration and NCCHD, 2005, 2006 and 2007

Year	NCCHD plurality	Registration plurality		
		Singleton	Multiple	Total
2005	Singleton	28,280	15	28,295
	Multiple	1	786	787
	Total	28,281	801	29,082
2006	Singleton	29,444	11	29,455
	Multiple	15	820	835
	Total	29,459	831	30,290
2007	Singleton	29,958	34	29,992
	Multiple	5	898	903
	Total	29,963	932	30,895

Source: Registration and NCCHD

Discordance in live or still birth status

For linked births in 2005 and 2006, there was 100 per cent concordance between live or still birth status recorded in birth registration and NCCHD data. For 2007, less than 0.1 per cent of records were discordant, as shown in Table 3.

Table 3 Comparison of birth status for linked births between registration and NCCHD, 2005, 2006 and 2007

Year	NCCHD birth status	Registration birth status			
		Live birth	Still birth	Total	Percentage
2005	Live birth	28,950	0	28,950	99.6
	Still birth	0	132	132	0.4
	Total	28,950	132	29,082	
2006	Live birth	30,154	0	30,154	99.6
	Still birth	0	136	136	0.4
	Total	30,154	136	30,290	
2007	Live birth	30,745	3	30,748	99.5
	Still birth	1	146	147	0.5
	Total	30,746	149	30,895	

Source: Registration and NCCHD

Discordance in babies' sex

The sex of the baby recorded at birth registration was compared with NCCHD in the linked dataset. For the three years, concordance of 99.9 per cent was observed on those records where the sex was stated in both the data sources (Table 4).

Table 4 Comparison of sex for linked live births between registration and NCCHD, 2005, 2006, 2007

NCCHD		Registration			
Year		Male	Female	Total	Percentage
2005	Male	14,876	16	14,892	51.4
	Female	17	14,021	14,038	48.5
	Not stated	12	8	20	<0.1
	Total	14,905	14,045	28,950	
2006	Male	15,439	15	15,454	51.2
	Female	14	14,674	14,688	48.7
	Not stated	6	6	12	<0.1
	Total	15,459	14,695	30,154	
2007	Male	15,737	27	15,764	51.3
	Female	16	14,930	14,946	48.6
	Not stated	16	20	36	0.1
	Total	15,769	14,977	30,746	

Source: Registration and NCCHD

Discordance in birth weight

Birth weights were grouped into 500 gram groups for comparison between linked records in birth registration and NCCHD. Where birth weight was stated, there was 99.3 per cent concordance between the two data sources in 2005 (Table 5a). Birth weight was missing from less than 0.1 per cent of NCCHD and birth registration records. In 2006 and 2007, 99.5 per cent of records were concordant. Missing birth weight data in NCCHD accounted for under 0.1 per cent of records in 2006 and 0.1 per cent of records in 2007.

Table 5a Comparison of birth weight distribution for linked births between registration and NCCHD, 2005

NCCHD birth weight (grams)	Birth registration birth weight (grams)													
	under 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
under 500	18	2	0	3	5	22	47	26	10	5	0	0	0	138
500–999	0	154	1	0	0	0	0	0	0	0	0	1	0	156
1000–1499	0	1	204	1	0	0	0	0	1	0	0	0	0	207
1500–1999	0	0	1	442	1	0	0	0	0	0	0	0	0	444
2000–2499	0	0	1	2	1,223	0	1	0	1	1	0	0	1	1,230
2500–2999	0	1	1	1	3	4,700	6	3	1	0	1	0	1	4,718
3000–3499	0	0	0	0	7	1	10,269	10	1	0	0	0	1	10,289
3500–3999	0	0	1	1	2	6	13	8,458	1	0	0	0	1	8,483
4000–4499	1	0	0	0	0	0	0	0	2,867	2	0	0	1	2,871
4500–4999	0	0	0	0	0	0	0	0	2	481	0	0	0	483
5000–5499	0	0	0	0	0	0	0	0	0	0	43	0	0	43
5500 and over	0	0	0	0	0	2	1	2	0	0	0	4	0	9
Not stated	1	0	0	1	0	1	3	4	1	0	0	0	0	11
Total	20	158	209	451	1,241	4,732	10,340	8,503	2,885	489	44	5	5	29,082

Source: Registration and NCCHD

Discordance in gestational age

Gestational ages recorded in NN4B and NCCHD were compared for all linked births. In 2005, 96 per cent of records with gestational age information were concordant between the two data sources as shown in Table 6a. Similar percentages were obtained for 2006 and 2007. Gestational ages were missing for between 0.4 per cent and 0.5 per cent of records in NCCHD over the three years.

Gestational ages differed by one week in around 2.3 per cent of records of births in 2005 and 2006. This dropped to 1.7 per cent of births in 2007. Gestational ages differed by two weeks or more in approximately 1.5 per cent of records in 2005. This was 1.8 per cent and 1.3 per cent in 2006 and 2007 respectively.

Table 6a Comparison of gestational age for linked births between NN4B and NCCHD, 2005

NCCHD Gestational age (weeks)	NN4B Gestational age (weeks)																							
	under 24	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43 and over	Not stated	Total	
under 24	15	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1	0	0	0	0	0	18	
24	1	28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	29	
25	0	0	30	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	31	
26	0	0	0	49	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	50	
27	0	0	0	1	40	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	41	
28	0	1	0	1	0	35	1	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	40	
29	0	0	0	0	0	2	50	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	54	
30	0	0	0	0	0	1	0	76	0	0	0	1	0	0	0	0	1	0	0	0	0	0	79	
31	0	0	0	0	0	0	0	2	90	2	0	0	0	1	0	1	0	0	0	0	0	0	96	
32	0	0	0	0	0	0	0	0	2	125	0	0	1	0	0	1	1	0	0	0	0	0	130	
33	0	0	0	0	0	0	0	0	0	0	152	0	1	0	0	4	0	0	0	0	1	0	158	
34	0	0	0	0	0	0	0	0	1	1	3	275	3	0	2	2	0	1	1	0	0	0	289	
35	0	0	1	0	0	0	0	0	0	0	0	4	391	4	2	4	4	2	3	0	0	0	415	
36	1	0	0	0	0	0	0	0	0	0	1	1	3	708	10	8	5	3	4	0	0	0	744	
37	0	0	0	0	0	0	0	0	0	2	0	1	2	6	1,403	15	7	8	12	1	0	0	1,457	
38	0	0	0	0	0	2	0	0	0	0	0	2	5	2	19	3,428	53	20	14	4	0	1	3,551	
39	1	0	0	0	0	0	1	0	0	0	0	0	1	1	11	50	5,801	120	22	2	4	0	6,014	
40	0	0	0	0	0	0	0	1	0	0	0	0	1	3	3	18	70	7,986	83	21	0	3	8,191	
41	1	0	0	0	0	0	0	0	0	0	0	0	0	1	4	9	24	91	5,899	44	3	3	6,079	
42	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	6	8	36	61	1,186	0	2	1,300	
43	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	7	10	24	17	53	0	114	
44 and over	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	3	16	20	6	0	20	0	68
Not stated	0	0	0	1	0	0	2	0	0	0	0	1	2	2	5	11	21	40	21	1	0	1	26	134
Total	19	29	31	53	41	40	54	79	93	130	156	287	411	728	1,461	3,562	6,008	8,334	6,164	1,282	61	30	29	29,082

Source: NHS Numbers for Babies and NCCHD

Discordance in ethnicity

The baby's ethnicities recorded on NN4B and NCCHD were compared for the linked records. Over 99 per cent of stated ethnic categories were concordant between the two datasets for all the years (Table 7a). For the three years, around 91 per cent to 92 per cent of babies with stated ethnicity belonged to the White group. In NCCHD, 13 per cent of records had no ethnicity recorded in 2005 but this fell to 6.8 per cent in 2006 and 1.3 per cent in 2007.

Table 7a Comparison of baby's ethnicity from NN4B and NCCHD for linked births, 2005

NCCHD Baby's ethnicity	NN4B Baby's ethnicity										Total
	White	Indian	Pakistani	Bangladeshi	Black African	Black Caribbean	Black other	Chinese	Any other ethnic group	Not stated	
White	23,143	1	4	0	0	0	0	0	21	13	23,182
Indian	0	245	0	0	0	0	0	0	1	0	246
Pakistani	3	1	234	2	0	0	1	0	1	0	242
Bangladeshi	1	1	2	122	0	0	0	0	0	0	126
Black African	2	0	0	0	153	0	2	0	1	1	159
Black Caribbean	0	0	0	0	0	13	0	0	3	0	16
Black other	4	0	0	0	0	0	36	0	0	0	40
Chinese	1	0	0	0	0	0	0	35	0	0	36
Any other ethnic group	75	10	2	4	1	2	2	31	992	2	1,121
Not stated	3,354	29	13	51	22	0	5	10	130	300	3,914
Total	26,583	287	255	179	176	15	46	76	1,149	316	29,082

Source: NHS Numbers for Babies and NCCHD

Discussion

The linkage process involved linking birth registration/NN4B records to NCCHD records, which were further linked to PEDW, to obtain key data items from maternity hospital discharge records. The main purpose of using the two Welsh data sources was that the maternity tails on PEDW are very incomplete (Welsh Government, 2012), but they contain important clinical data items which are not included in the child health system. The aim therefore was to extract the best available information for common data items and to include unique data items from each data source to compile a good quality dataset. Because only data from NCCHD were provided for common data items, it was not possible to compare these data fields with those also recorded in PEDW. All the key data items, for which the results have been reported in this paper, came from NCCHD.

The linkage algorithms were similar to those used in linking registration/NN4B to Maternity HES (Dattani *et al*, 2011). Over 97 per cent of the linked registration/NN4B/NCCHD records were linked to PEDW using the mother's NHS number. The mother's NHS number was missing from over 30 per cent of birth records in the NN4B dataset in 2005 and 2006 and ONS did not provide this information for 2007 births. Therefore, the registration/NN4B linked dataset was first linked to NCCHD using the baby's NHS number to extract the maximum number of baby records which had their mother's NHS number. The next step involved using the mother's NHS number from NCCHD to link to PEDW to add delivery records. Baby fields in PEDW were not included because so many were incomplete. In the final linked dataset, some fields were derived from NCCHD and some from PEDW. A number of fields common to NCCHD and PEDW, such as birth weight and gestational age, were obtained from NCCHD as it is known to be of better quality than the maternity tail in PEDW at an all-Wales level. The NCCHD has previously been linked to PEDW for births occurring in 2008, to link birth data from the two main sources of birth and delivery information (Welsh Government, 2010). Linkage of NCCHD and PEDW to birth registration and NN4B linked data has added important data about care at birth. Additionally, the linkage is a good way of cross-validating the values of the common data items in the two datasets.

The linkage rate was approximately 92 per cent. This was similar to the rate achieved when linking registration/NN4B records for England to Maternity HES but the data in all the fields of interest in NCCHD were much more complete than in the corresponding fields in Maternity HES. Information about birth weight, gestational age, and live/ still birth status were missing from over 20 per cent of records in Maternity HES (Dattani *et al*, 2011). For Wales, PEDW data are equivalent to Maternity HES but their completeness could not be assessed and compared for the reasons stated earlier, since the key data items were not obtained from PEDW.

Over the three years, between 2.7 per cent and 2.9 per cent of records in birth registration and NCCHD were for multiple births. Multiple births were included in the data quality checks as information was adequately recorded for each birth within the sets of multiple births. This was unlike the situation with the Maternity HES linkage. The information was often recorded for the first baby only, while information for the subsequent babies from the multiple birth was either missing or a repeat of the information about the first baby (Dattani *et al*, 2011).

Gestational age data are available from the NN4B dataset for all births whereas at birth registration, gestational age is recorded only for still births. Therefore, NN4B data were used for comparison purposes. Gestational age can be assessed at second trimester ultrasounds which are now routinely done. The other usual method is the calculation of gestational age in weeks of time from the first day

of the woman's last menstrual period (LMP). Gestational age distributions have been shown to differ depending on method used to assess gestational age (Mongelli and Gardosi, 1996; Savitz *et al*, 2002; Yang *et al*, 2002). Gestational ages recorded in the two data sources were highly concordant and differed by one week or more on just over 4 per cent of records in 2005 and under 4 per cent in subsequent years. A 100 per cent concordance would be expected, as from 2002 onwards, records in the community child health databases and NN4B records have been derived from a common source. Nonetheless, local health boards can amend or add to the records once they have received the data.

Ethnicity was categorised into 11 groups (including not stated), to have consistent coding for ethnicity in both the datasets. In NCCHD, prior to 2001, ethnicity was recorded using a 10-group classification and from 2001 onwards it was replaced by the 2001 Census classification. Both these options are included in the input options on the NCCHD, however some users still select the original group codes. There is some uncertainty about the recording of the ethnic group in the NN4B dataset - this is because the system requests information about the ethnic category of the baby as defined by the mother using the 2001 Census categories (Birth Notification Dataset, 2001). It is unclear to what extent mothers or health professionals define this ethnic category. In addition, although the ethnic group of the baby is requested in NN4B, there might be occasions when the mother's ethnic group is not actually recorded. Over 99 per cent of records with stated ethnicity were concordant between NN4B and NCCHD but this may simply be a consequence of the way data flow between these two systems.

Conclusion

Although the linkage rate was similar to that obtained in linking registration/NN4B linked data to Maternity HES in England, the high completeness and the consistency of the key data items between the datasets suggest that the key information recorded in NCCHD is reliable. NCCHD linked to PEDW is a reliable source of data and can be used to look at birth data for Wales without the need for linkage to data from birth registration or NN4B. Nevertheless, important data items such as the parents' countries of birth and socio-economic status are recorded only at birth registration. This means that linkage to this dataset is needed to enable analyses of birth outcomes by ethnicity, socio-economic status and parents' countries of birth.

Acknowledgement

This work forms part of the 'Linkage, analyses and dissemination of national birth and maternity data for England and Wales' project, funded by Medical Research Council as part of the Joint Wellcome Research Councils Electronic patient data linkage initiative. We would like to thank Russell Brown, from NHS Wales Informatics Service (NWIS), for linking the datasets; Julie Messer at ONS for providing the birth registration/NHS Numbers for Babies linked data to NWIS to link to PEDW and NCCHD, making the linked data accessible in the VML system; Vanessa Fearn for releasing the outputs. Collaborators in the original National Gestational Age project included, in addition to the authors, Lesz Lancucki, formerly Maternity Hospital Episodes Statistics, Community Health Statistics and Surveys, NHS Informational Centre and Tony Couch, formerly Head of Information Products, Health Solutions Wales who we like to thank for their help in the earlier stage of the project. We are grateful to Martin Ward Platt, Clinical Director, Regional and Maternity Surveys Office, North East Region for his help in this project.

References

- Brocklehurst P (1999). Infection and preterm delivery. *British Medical Journal*, 318, pp. 548–549.
- Confidential Enquiry into Maternal and Child Health (2004). Still birth, neonatal and postneonatal mortality 2000–02, England, Wales and Northern Ireland.
- Connecting for Health (2001) *Birth Notification Data Set*. Available from: www.connectingforhealth.nhs.uk/industry/docs/nn4b/nn4bdataset.pdf [Accessed 10/01/2012].
- Dattani N, Datta-Nemdharry P and Macfarlane A (2011). Linking maternity data for England, 2005–06: methods and data quality. *Health Statistics Quarterly*, 49, pp. 53–80.
- Dattani N, Datta-Nemdharry P and Macfarlane A (2012). Linking maternity data for England, 2007: methods and data quality. *Health Statistics Quarterly*, 53, pp. 4–21.
- Euro-Peristat (n.d.) *European perinatal health report 2008*. Available from: www.europeristat.com [Accessed 29.10.10].
- 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, pp. 25–33.
- Mongelli M and Gardosi J (1996). ‘Gestation-adjusted projection of estimated fetal weight’, *Acta Obstetrica et Gynecologica Scandinavica*, 75(1), pp. 28–31.
- Moser K and Hilder L (2008). Assessing quality of NHS Numbers for Babies data and providing gestational age statistics. *Health Statistics Quarterly*, 37, pp. 15–23.
- National Health Service Act 1977, S124 (4–7).
- NHS Wales Informatics Service (2009.) *What is PEDW?* Available from: www.infoandstats.wales.nhs.uk/page.cfm?pid=41009&orgid=869 [Accessed 03/02/2012].
- Office for National Statistics (2010) *Births Tables: Metadata 2010*.
- Office for National Statistics (2011) *Statistical Bulletin: Quality of ethnicity and gestation data subnationally for births and infant deaths in England and Wales, 2005-2008*. Available from: www.ons.gov.uk/ons/rel/child-health/quality-of-ethnicity-and-gestation-data-subnationally-for-births-and-infant-deaths-in-england-and-wales/2005-08/stb.html [Accessed 29/04/2012].
- Savitz D A, Terry J W Jr., Dole N, Thorp J M, Siega-Riz A M and Herring A H (2002) ‘Comparison of pregnancy dating by last menstrual period, ultrasound scanning and their combination’, *American Journal of Obstetrics and Gynaecology*, 187(6), pp. 1660–1666.
- Welsh Government (2010) *Adding value to the analysis of Welsh birth data, 2008*. Available from: www.wales.gov.uk/docs/statistics/2010/101215addingvalueen.pdf [Accessed 03/02/2012].
- Welsh Government (2012) *Maternity Statistics, Wales: Method of Delivery, 2001–2011*. Available from: www.wales.gov.uk/docs/statistics/2012/120131sdr132012en.pdf [Accessed 03/02/2012].

Yang H, Kramer M S, Platt R W, Blondel B, Breart G, Morin I, Wilkins R and Usher R (2002) 'How does early ultrasound scan estimation of gestational age lead to higher rates of preterm birth?', *American Journal of Obstetrics and Gynaecology*, 186(3), pp. 433–437.

Appendix A

For linkage of registration and NHS numbers for babies (NN4B) linked data to National Community Child Health Database for Wales (NCCHD) and Patient Episode Database Wales (PEDW) a file consisting of the small subset of data items shown in Table A1, including the mother's and baby's NHS number, date of birth of mother, baby's date of birth, their postcode and a unique ID compiled by ONS, was sent to NWIS. Indirect identifiers were used for linkage of records where the mother's or baby's NHS number was missing. This included various combinations of mother's date of birth and post code.

NWIS did the linkage in two phases: Firstly the linked registration/NN4B data were linked to NCCHD using the baby's NHS number. The next step involved linking the data created in the first phase to PEDW. This included using the mother's NHS number from NCCHD and combinations of the mother's date of birth and postcode either with operation codes R14–R27 or diagnostic codes O00–O99. A variable indicating the stage of the algorithm used for linking the linked registration/NN4B/NCCHD data to PEDW was provided by NWIS.

Table A1 Number and percentage of registration/NN4B/NCCHD records that were linked to PEDW records by algorithm, 2005–07

Stage	Variables used	Records linked					
		2005		2006		2007	
		Number	%	Number	%	Number	%
Linkage to PEDW							
1	Mother's NHS number with operation code R14-R27	28,363	97.5	29,659	97.9	30,129	97.5
2	DOB + Postcode with operation code R14-R27	461	1.6	379	1.3	295	1.0
3	DOB + Partial Postcode with operation code R14-R27	33	0.1	29	0.1	25	0.1
4	Mother's NHS number with diagnostic code O00-O99	219	0.8	217	0.7	443	1.4
5	DOB + Postcode with diagnostic code O00-O99	5	0.0	5	0.0	3	0.0
6	DOB + Partial Postcode with diagnostic code O00-O99	1	0.0	1	0.0	0	0.0

1. For 2005 there were 2676 (8.4%) registration/NN4B records that did not match.
2. For 2006 there were 2508 (7.7%) registration/NN4B records that did not match.
3. For 2007 there were 2697 (8%) registration/NN4B records that did not match.

Source: Registration, NHS Numbers for babies, NCCHD, PEDW