Millennium Cohort Study:

Birth Registration and Maternity Hospital Episode Statistics Linkage

A Guide to the SPSS Dataset

February 2006

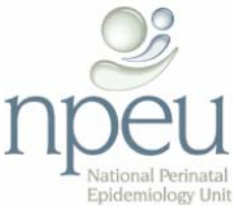
Carole Dezateux, Christine Hockley, Jon Johnson, Heather Joshi, Maria Quigley, Rachel Rosenberg

Centre for Longitudinal Studies

Bedford Group for Lifecourse & Statistical Studies

Institute of Education, University of London

In collaboration with


npeu
National Perinatal
Epidemiology Unit

ICH LOGO


E·S·R·C
ECONOMIC
& SOCIAL
RESEARCH
COUNCIL

Contents

Background

Data Sources

Millennium Cohort Study

The Millennium Cohort Study (MCS) is a longitudinal observational study of nearly 19,000 babies born in the UK between September 2000 and January 2002 [1-2]. Parents of the babies were recruited from Child Benefit Records after being given a chance to opt out by post, telephone or on the doorstep. Parents were interviewed for the first time (sweep 1) when the baby was about 9 months old, and information was collected on social, economic, educational and health factors. The health questions included topics such as the mother's health during pregnancy, birth and delivery, and the baby's health immediately after birth. However, the time constraints on the MCS interview limited the number of health-related questions included in the questionnaire. Furthermore, mothers may not have been able to recall accurately some of the information on pregnancy and delivery 9 months later. Mothers responding to the 9 month interview were accordingly asked to give written consent to the following records being added to the survey

Birth registration records

Personal and demographic information about baby and parents at birth registration is available from the Office for National Statistics (ONS) in England and Wales, the General Register Office (GRO(S)) in Scotland and the Northern Ireland Statistics and Research Agency (NISRA), Demography and Methodology Branch in Northern Ireland.

Centrally collected hospital records

The data in hospital records in all countries are held in the *general record.* This contains information about the mother's stay in hospital, including details of any operations performed or any diagnoses that were made during the hospital stay, the length of stay in hospital, etc. England, Wales and Scotland also hold extra information relating to the birth of the baby, such as the birth weight, delivery method, etc. in the *tail record.* Not all hospitals in England and Wales are able to submit this additional information. Northern Ireland does not hold any maternity 'tail' variables at all.

Health information is held centrally by the DH in England, by Health Solutions Wales (HSW) of the NHS in Wales and the Information and Statistics Division (ISD) of the NHS in Scotland. In Northern Ireland, general health records are available centrally from the Department of Health, Social Services and Public Safety (DHSSPS) and maternity information from the four area (Northern, Southern, Eastern and Western) Health and Social Services (HSS) boards in Northern Ireland. In all four countries, trusts or hospitals are required to submit data held on their admissions systems to the relevant country's health services department.

Record Linkage methods.

Birth registration records

The identifiers necessary to link MCS cases to birth registration data in each country were identified (Figure 1. Stage. 1 linkage variables). These data items were selected from the

main interview data for each consenting cohort member and were sent to the holders of birth registration information for each country.

In addition to obtaining additional variables, the birth registration linkage yielded the postcode at birth of the cohort member for more accurate matching against hospital episodes; the MCS had data on postcode at interview rather than birth. Furthermore, the opportunity to flag and trace all cohort members who had consented was taken for future use. The NHS number for the baby(ies) was obtained as part of the flagging and tracing exercise and this was used to match against the birth registration data. In addition, coding of mother's and father's occupation (SOC 2000) was undertaken for England, Wales and Northern Ireland at the same time. Coding was performed by the holders of the birth registration data in these countries. Records in Scotland were already coded for occupation so this information was included in the data returned. Details of the variables gained by the linkage are shown in Figure 1 (Stage 1: Data items gained through linkage).

Centrally collected hospital records

The variables needed for linkage to hospital records (Figure1. Stage 2. Linkage variables) were selected from the information collected at the MCS interview and that received from birth registration linkage. These data were split into four different datasets, one for each UK country, and only those cases for which consent for hospital episode linkage had been obtained were included. For the 19 cases who had given consent to hospital episode records, but not birth registration linkage, the postcode at interview was used as a proxy for postcode at birth, except for those who had moved since birth (approx.12%), where postcode at birth was substituted if reported in the interview.

The variables used for record linkage varied by country (Table 1). In England, Wales and Northern Ireland, record linkage was performed using a deterministic method of matching whereby a match was only accepted if there was complete agreement on a given combination of common variables. Hence, it is theoretically possible that false matches will occur if records belonging to two different individuals have identical matching variables.

Matching in England used differing combinations of the whole and partial variables provided using a hierarchical approach [3]. This catered for missing and incomplete data yet still allowed for as good a match as possible. Match types 1 to 5 catered for different levels of completion for those with 'tails' or *partial 'tails'* and match type 6 catered for those without 'tail' information. Thus, the strength of the match diminished with each type, with type 1 achieving a highly probably match and type 6 achieving a match that could not be accepted as true without checking other variables such as date of admission as a surrogate for baby's date of birth. Table 1 (below) and Appendix 2, Table 2 show the variables and match types that were used in this hierarchical matching process.

The only common variables between the MCS database and the databases in Wales and Northern Ireland were postcode, mother's date of birth, hospital of birth and (in Wales only) mother's and baby's names. The main matching variables were postcode and mother's date of birth. Hospital of birth was used to verify matches achieved. There is little information available in Northern Ireland concerning the baby and in Wales the submission of 'tail' information is very low (22%). However, the baby's date of birth provided from MCS was compared with the mother's operation date (Northern Ireland) or the mother's admission date (Wales).

Scotland adopted a probabilistic method of record linkage entailing comparisons with all possible pairs of records and making a decision as to whether they belonged to the same individual. Every time an item of identifying information was the same on the two records, the probability that they applied to the same person increased. A probability matching algorithm was created which took into account the discrepancies and missing values that occur in most large datasets.

Table 1 Variables used for linking MCS data to hospital records by country.

Variable	England*	Wales	Scotland	N Ireland
Baby's date of birth	✓ 1,2,3,4,5		✓	
Mother's date of birth	✓ 1,2,3,4,5,6	✓	✓	✓
Baby's sex	✓ 1,2,3		✓	
Baby's birth weight	✓ 2,3, 5		✓	
Postcode at birth	✓ 1,2,3,4,5,6	✓	✓	✓
Mother's name		✓	✓	
Baby's name		✓	✓	
Hospital of birth	✓ 6	✓	✓	✓

* In England, the hierarchical match type which required this matching variable is also listed
Note that some variables were used completely (e.g. mother's date of birth) for some match types and partially (e.g. mother's month and year of birth) for other match types.

Details of the variables gained by the linkage are detailed in Appendix 3, Table 1: Hospital Records: Completion of variables in the general record and the 'tail' by country.

Singletons and multiples

Matching was done separately for singletons and multiples. This was due to the complexity of the data on multiples who have individual "baby data items" (e.g. birth weight), shared "mother data items" (e.g. mother's date of birth) and multiple-specific "perinatal data items" (e.g. birth order). Hence, most results are reported separately in singletons and multiples.

Births at home or in private hospitals

The vast majority of UK deliveries occur in NHS hospitals. In 2001-2002, about 2% of deliveries in England occurred at home and 0.5% occurred in private hospitals [4]. Women who give birth at home or in a private hospital will not always have details of their delivery submitted to central hospital records. For example, in 1998-99, only 12% of home births in England had a hospital record submitted to HES [4]. In the MCS, 345 women (1.86%) reported that they had had a home birth: 254 (2.20%) in England; 58 (2.10%) in Wales; 28 (1.20%) in Scotland; and 5 (0.26%) in Northern Ireland. We do not know what proportion of MCS babies were born in private hospitals. As a result, matching hospital records will only be found for small proportions of births that were at home or in private hospitals.

Success of matching

Birth Registration

In June 2003, CLS sent 12,096 records for birth registration linkage to ONS (for England and Wales), 2,087 records to GRO(S) (for Scotland), and 1,686 records to NISRA (for Northern Ireland). All matched data were received by NPEU by March 2004. In April/June 2005, a second batch was sent by CLS to all four countries (1,040 to England and Wales, 92 to

Scotland and 18 to Northern Ireland), and all matched data were received by NPEU by November 2005.

Table 2 shows the number of matched birth registration records that were found in each country. Matching was extremely successful in all countries: of those who consented to record linkage, a matching record was found for 99% (16,840/17,019) of babies. This means that 89% (16,840/18,818) of MCS babies have a matching birth registration record (those who gave consent and had a matching record). This figure ranges from 83% in Northern Ireland to 92% in Scotland.

Table 2 Birth registration: number of matched records found by country and multiplicity.

	Total MCS families	Total MCS Babies	No. (%) who gave consent for birth egistration	No. (%) matched[1]	% matched[2]
England	**11,532**	**11,694**	**10,542 (90%)**	**10,474 (99%)**	**90%**
Singletons		11,375	10,250 (90%)	10,182 (99%)	90%
Multiples		320	292 (91%)	292 (100%)	91%
Wales	**2,761**	**2,799**	**2,594 (93%)**	**2,578 (99%)**	**92%**
Singletons		2,725	2,527 (93%)	2,514 (99%)	92%
Multiples		74	67 (91%)	64 (96%)	86%
Scotland	**2,336**	**2,370**	**2,179 (92%)**	**2,173 (100%)**	**92%**
Singletons		2,305	2,122 (92%)	2,116 (100%)	92%
Multiples		65	57 (88%)	57 (100%)	88%
Northern Ireland	**1,923**	**1,955**	**1,704 (87%)**	**1,615 (95%)**	**83%**
Singletons		1,892	1,655 (87%)	1,569 (95%)	83%
Multiples		63	49 (78%)	46 (94%)	73%
Total babies	**18,552**	**18,818**	**17,019 (90%)**	**16,840 (99%)**	**89%**

1 using the number who gave consent as the denominator
2 using the total MCS babies as the denominator

Hospital Records

In October 2003, CLS sent 9,623 records for hospital episode linkage to HES (for England), 2,347 records to HSW (for Wales), 2,085 records to ISD (for Scotland), and 1,676 records to DHSSPS (for Northern Ireland). All matched data were received by NPEU by March 2004. In June 2005, a second batch of data was sent by CLS to all four countries (795 records to HES, 257 records to HSW, 75 records to ISD and 35 records to DHSSPS). All matched data were received by NPEU by December 2005.

As described previously, the data in hospital records in all countries are held in the *general record* which contains "general" information about the mother's stay in hospital. England, Wales and Scotland also hold extra perinatal information in the *tail record.* Northern Ireland does not contain any maternity 'tail' variables.

For England, Wales and Scotland the matching variables are contained in the 'tail' with the exception of mother's date of birth which is in the *general record*. Some hospitals do not submit 'tail' information at all, others submit 'tails' with incomplete variables. A successful match depends on accurate data being entered into the matching variables. Results showing completion of the matching variables in the hospital records for England are shown in **Appendix 2** (Table 1). For those with a matched record in England, completion in HES was at least 85% for all matching variables except whole postcode (67%), baby's date of birth (72%) and baby's birth weight (73%). In Northern Ireland, matching was done using mother's date of birth, postcode and hospital of birth.

Table 3 shows the success of matching MCS data to hospital record data in all four countries. A matching hospital record was found for 83% of those who gave consent. Therefore, 75% (14,079/18,818) of MCS babies have a matching hospital record (those who gave consent and had a matching record). This figure ranges from 58% in Northern Ireland to 85% in Scotland and Wales.

There was substantial variation in matching rates by country, with the lowest rate occurring in Northern Ireland (66%) and the highest in Scotland (93%). In England, Scotland and Northern Ireland, matching rates varied slightly between singletons and multiples, whereas in Wales, the matching rate was substantially lower in the small number of multiples (44/59=75%) than in singletons (91%).

Table 3 Hospital records: number of matched records found by country and multiplicity.

	Total MCS families	**Total MCS babies**	**No. (%) who gave consent for health records**	**No. (%) matched[1]**	**% matched[2]**
England	**11,532**	**11,694**	**10,418 (89%)**	**8,689 (83%)**	**74%**
Singletons		11,375	10,158 (89%)	8,458 (83%)	74%
Multiples		320	260 (81%)	231 (89%)	72%
Wales	**2,761**	**2,799**	**2,604 (93%)**	**2,370 (91%)**	**85%**
Singletons		2,725	2,545 (93%)	2,326 (91%)	85%
Multiples		74	59 (80%)	44 (75%)	59%

Scotland	**2,336**	**2,370**	**2,160 (91%)**	**2,004 (93%)**	**85%**
Singletons		2,305	2,103 (91%)	1,978 (94%)	86%
Multiples		65	57 (88%)	55 (96%)	85%
Northern Ireland	**1,923**	**1,955**	**1,707 (87%)**	**1,133 (66%)**	**58%**
Singletons		1,892	1,655 (87%)	1,103 (67%)	58%
Multiples		63	49 (78%)	30 (61%)	48%
Total babies	**18,552**	**18,818**	**16,889 (90%)**	**14,196 (84%)**	**75%**

1 using the number who gave consent as the denominator
2 using the total MCS babies as the denominator

Table 4 shows the proportion of matched records with a 'tail'. The four countries varied markedly in terms of the proportion of matched records that had a 'tail': in Northern Ireland no 'tail' information was present, in Wales, only 22% of matched records had a 'tail' (the remaining 78% had a general record only), whereas in England 80% had a 'tail', and in Scotland 100% had a 'tail'. However, it should be noted that some of these 'tails' were not well completed (see below and Appendix 3). Among the multiple births, all matched records in England and Scotland had 'tails' whereas none of the matched records in Wales and Northern Ireland had 'tails'. In general, the proportion of matched records with a 'tail' was higher than expected from the feasibility study undertaken by NPEU in 2002..

More detailed results showing the success of matching according to particular criteria are shown in **Appendix 2**. For England, the success of matching by *match type* is shown in **Appendix 2** (Table 2). For Wales and Northern Ireland, the proportion of records satisfying particular matching criteria are shown in **Appendix 2** (Table 4 for Wales and Table 5 for Northern Ireland). No such table is presented for Scotland because probabilistic matching was employed.

Table 4 Hospital records: proportion of matched records which have a 'tail' by country and multiplicity.

Country	Singletons			Multiples			Overall Total
	'Tail[1]'	No 'tail'	Total	'Tail[1]'	No 'tail'	Total	
	N (%)	N (%)	N	N (%)	N (%)	N	N
England	6,797 (80%)	1,661 (20%)	**8458**	231 (100%)	0 (0%)	**231**	**8,689**
Wales	515 (22%)	1811 (78%)	**2326**	0 (0%)	44 (100%)	**44**	**2,370**
Scotland	1978 (100%)	0 (0%)	**1978**	55 (100%)	0 (0%)	**55**	**2,033**
Northern Ireland	0 (0%)	1103 (100%)	**1103**	0 (0%)	30 (100%)	**30**	**1,133**
Total	**9,290 (67%)**	**4,575 (33%)**	**13,865**	**286 (79%)**	**74 (21%)**	**360**	**14,225**

1 Note that where a tail is indicated, the tail variables are not necessarily complete.

Deposited Datasets

Structure of the data

The Scottish birth registration data and the English hospital data contained many more fields than the other two countries, and so where a field was not present a dummy variable has been inserted.

Each country has a birth registration and a hospital data set with the same number of variables in the identical order as for the other countries. The country of birth is not necessarily the same as the country at MCS1 interview.

Birth registration Data

Country	Filename
England	br_eng.sav
Wales	br_wales.sav
Scotland	br_scot.sav
Northern Ireland	br_ni.sav

Number of variables: 17

Variable names and labels for each country

Name				Label
England	**N.Ireland**	**Scotland**	**Wales**	
famsrno	famsrno	famsrno	famsrno	BR: Family serial number
cnum	cnum	cnum	cnum	BR: MCS child number of baby
mocccd	mocccde	mothoccd	mocccd	BR: Mother's SOC 2000 occupation code
mcbcode	**mcbcode**	mcbcode	mcbcode	BR: Mother's country of birth code
mothemp	**mothemp**	mothemp	**mothemp**	BR: Mother's employment status and code
mothsoc	**mothsoc**	mothsoc	**mothsoc**	BR: Mother's social class
mnssec	**mnssec**	mnssec	**mnssec**	BR: Mother's NSSEC category
focccd	focccde	focccd	focccd	BR: Father's SOC 2000 occupation code
fcbcode	**fcbcode**	fcbcode	**fcbcode**	BR: Father's country of birth
fathemp	**fathemp**	fathemp	**fathemp**	BR: Father's employment status and code
fsocc	**fsocc**	fsocc	**fsocc**	BR: Father's social class
fnssec	**fnssec**	fnssec	**fnssec**	BR: Father's NSSEC category
bthwto	**bthwto**	**bthwto**	bthwto	BR: Birth weight (grams) of baby
bsex	babysex	sex	bsex	BR: Sex of baby
totprevp	**totprevp**	totprevb	**totprevp**	BR: Number of previous pregnancies
flag	flag	flag	flag	BR: Inconsistent data when compared with HES/MCS
country	country	country	country	BR: Country

N.B. Data which was not collected in a particular country are highlighted in **bold.**

Country of Birth Coding
The coding frame used for Country is ISO-3166 a full list is available from the ISO website at http://www.iso.org/iso/en/prods-services/iso3166ma/02iso-3166-code-lists/list-en1.html.

Occupation coding
The coding frame is that of ONS full details are available from the ONS website at:
http://www.statistics.gov.uk/methods_quality/ns_sec/soc2000.asp

Hospital Records

Country	**Filename**
England	hosp_eng.sav
Wales	hosp _wales.sav
Scotland	hosp _scot.sav
Northern Ireland	hosp _ni.sav

Number of variables: 65

Variable names and labels for each country

Name				**Label**
England	**N.Ireland**	**Scotland**	**Wales**	
famsrno	famsrno	famsrno	famsrno	Family Serial Number
cnum	cnum	cnum	cnum	MCS Child Number
admimon	admon	admon	admimon	Month of admission
admiyr	admyr	admyr	admiyr	Year of admission
admimeth	admit_me	transfrm	admimeth	Method of admission
admisorc	admit_so	**admisorc**	admisorc	Source of admission
classpat	category	patcat	classpat	Patient classification
disdest	disdest	transto	disdest	Destination on discharge
dismon	dismon	dismon	dismon	Month of discharge from hospital
disyr	disyr	disyr	disyr	Year of discharge from hospital
epidur	eps_dur	**epidur**	epidur	Episode duration
ethnos	**ethnos**	ethnic	ethnos	Ethnic origin
hrglate	hrg	**hrglate**	**hrglate**	Healthcare Resource Group code
mainspef	spec_cod	specty	mainspef	Main specialty
postdur	**postdur**	**postdur**	**postdur**	Post natal stay
diag1_3c	pdiag_3	mdiag_3	diag1_3	ICD10 3-char Diagnosis code 1
diag2_3c	secdi1_3	odiag1_3	diag2_3	ICD10 3-char Diagnosis code 2
diag3_3c	secdi2_3	odiag2_3	diag3_3	ICD10 3-char Diagnosis code 3
diag4_3c	secdi3_3	odiag3_3	diag4_3	ICD10 3-char Diagnosis code 4
diag5_3c	secdi4_3	odiag4_3	diag5_3	ICD10 3-char Diagnosis code 5
diag6_3c	secdi5_3	odiag5_3	diag6_3	ICD10 3-char Diagnosis code 6
diag7_3c	**secdi6_3**	**odiag6_3**	diag7_3	ICD10 3-char Diagnosis code 7
oper1_3c	pop_3	mainopa3	oper1_3	OPCS4 3-char Operation code 1
oper2_3c	secop1_3	mainopb3	oper2_3	OPCS4 3-char Operation code 2
oper3_3c	secop2_3	othop1a3	oper3_3	OPCS4 3-char Operation code 3
oper4_3c	secop3_3	othop1b3	oper4_3	OPCS4 3-char Operation code 4
oper5_3c	**secop4_3**	othop2a3	oper5_3	OPCS4 3-char Operation code 5
oper6_3c	**secop5_3**	othop2b3	oper6_3	OPCS4 3-char Operation code 6
oper7_3c	**secop6_3**	othop3a3	oper7_3	OPCS4 3-char Operation code 7
oper8_3c	**secop7_3**	othop3b3	oper8_3	OPCS4 3-char Operation code 8
preopdur	propstay	**preopdur**	**preopdur**	Pre operation duration
posopdur	**posopdur**	**posopdur**	**posopdur**	Post operation duration
resgorn	**resgor**	**resgor**	**resgor**	Government office region
startage	**hisage**	**startage**	**startage**	Mother's age at start of episode
tretspef	**tretspef**	signfac	tretspef	Treatment specialty
totcost	**totcost**	**totcost**	**totcost**	Notional cost of treatment (£)

treat	**treat**	**treat**	**treat**	Cost of treatment (£)
anasmon	**anasmon**	antemon	anasmon	Month of first antenatal assessment
anasyr	**anasyr**	anteyr	anasyr	Year of first antenatal assessment
antedur	**antedur**	**antedur**	**antedur**	Antenatal days of stay
biresus	**biresus**	resusc	biresus	Resuscitation method
birordr	**birordr**	**birordr**	birord	Birth order
birweit	**birweit**	bwt	bwt	Birth weight of baby
delchang	**delchan**	delchan	delchan	Change of delivery place
delinten	**delinten**	delplac	delintn	Delivery place (intended)
delmeth	**delmeth**	delmode	delmeth	Delivery method
delonse	**delonse**	**delonse**	delonse	Method to induce labour
delplac	**delplac**	**delplac2**	delplac	Delivery place (actual)
delposan	**delposan**	**delposn**	delposn	Anaesthetic given post-labour or delivery
delprean	**delprean**	anallab	delpren	Anaesthetic given during labour or delivery
delanal	**delanal**	delanal	**delanal**	Analgesia during delivery
labdur	**labdur**	**labdur**	**labdur**	Duration of labour
delstat	**delstat**	sendoc	delstat	Status of person conducting delivery
gestat	**gestat**	gestwks	gestat	Length of gestation
numbaby	**numbaby**	numbabs	numbaby	Number of babies
numpreg	**numpreg**	numpreg	numpreg	Number of previous pregnancies
sexbaby	**sexbaby**	bsex	sexbaby	Sex of baby
neocare	**neocare**	neocare	neocare	Neonatal level of care
match_type	goodmtch	goodmtchs	match_type	type of match with MCS
tailornot	**tailornot**	**tailornot**	tailornot	tail items present
numbflag	**numbflag**	**numbflag**	**numbflag**	Reliability flag for number of babies
liveflag	**liveflag**	**liveflag**	**liveflag**	Reliability flag for live birth
sexflag	**sexflag**	**sexflag**	**sexflag**	Reliability flag for sex of baby
flag	flag	flag	flag	inconsistent data when compared with BR/MCS
country	country	country	country	Country

N.B. Data which was not collected in a particular country are highlighted in **bold.**

Diagnosis Coding

Diagnosis has been coded to International Statistical Classification of Diseases and Related Health Problems 10th Revision (ICD-10), the coding frame is available on the World Health Organisation website at: http://www.who.int/classifications/apps/icd/icd10online/. Only the first three characters are available in this dataset.

Operation Coding

Operations were coded to The Office of Population Censuses and Surveys' Classification of Surgical Operations and Procedures – Fourth Revision (OPCS-4). Only the first three characters are available in this dataset.

Costs of treatment

This has been recoded to bands of £250 for notional cost of treatment and £100 for actual cost.

Un-deposited Data

Some data collected during the linkage exercise has not been deposited in this dataset for access under the standard End User Licence for disclosure reasons.

Birth Registration

Day of birth of cohort member
Day of birth of mother
Hospital

Hospital Records

Diagnosis coding ICD-10 – full four character code
Operation coding OPCS4 - full four character code
Day of episode start
Day of episode end
Health Authority

Accuracy of matching methods

Deterministic method of matching used for linking hospital records

In order to help determine whether the records matched under each match step were likely to be true matches, the probability of observing each match type purely by chance was estimated using the data for England. For example, in theory, there is a 50% chance that any two records would have the same sex and an 8.3% chance (1 in 12) that any two records would have the same month of birth. These probabilities were estimated for the matching variables individually (see **Appendix 1**, Table 1) and in combination. If two records were extremely unlikely to match by chance on the fields specified in the match type, then it was assumed that the records in this match type probably belong to the same person.

The probabilities for the different match types are shown in **Appendix 1** (Table 2). All of the probabilities shown here are small enough to rule out chance with the exception of match types 4a and 6 where, if postcode were not included, the probability is 0.28% in HES and 0.136% in MCS for match type 4a, and 0.02% in HES and 0.016% in MCS for match type 6. Once a postcode match is added to match types 4a and 6, we believe that these probabilities would be sufficiently small to make chance unlikely. While these probability estimates are crude (see the limitations expressed in **Appendix 1**), they give an indication as to whether each match type is likely to occur by chance. It is apparent that, the probabilities of two records matching purely by chance is extremely small for all match types.

These crude probability estimates for England may also be applied to Wales and Northern Ireland.

Probabilistic method of matching used in Scotland for linking hospital records

Each MCS record was compared with each hospital delivery record and a score was assigned when a match was identified between linking variables. The scores were weighted according to the probability that the pairs of records belong to the same individual. For example, an exact match on date of birth would create a high score, whereas a lower score would be assigned if there was a difference of one digit. Matches on common surnames, such as Smith, would have a lower score than a match on a more unusual name.

If several variables match between one pair of records, then the linkage score accumulates for that pair. Following a detailed examination of a sample of these pairs, a threshold was set whereby all linkages with a score higher than the threshold were deemed to be good links. A cautious approach was taken and the threshold was set quite high to ensure a high level of accuracy. This approach did not seem to adversely affect the quantity of records successfully linked. It was possible to match 93% of the MCS records with mother's delivery records.

Completion of variables gained through matching

Birth Registration

Table 5 shows the completion of the variables gained through matching MCS data to birth registration. Note that the birth registration records contain additional variables to those listed in Table 5 (e.g. marital status, joint registration, etc) but these were not sought as part of the record linkage project. The number of additional variables gained varies by country, but is highest in Scotland. The only additional variables available in all four countries are mother's and father's occupation and these have been coded to SOC (Socio-economic classification) 1990 or SOC 2000. Mother's occupation was more complete for multiples in all countries than for singletons, ranging from 84% for multiples in Scotland to 88% for multiples in Wales. Similarly, the completion of father's occupation ranged from 87% for singleton babies in Wales to 100% for multiples in Scotland. Birth weight is available from birth registration for all MCS babies in England and Wales but not for those born in Scotland and Northern Ireland. Mother's place of birth was available for all babies in England and Wales and has been used to derive a country of birth code which is equivalent to that provided by Scotland. Place of birth is usually reported as town and county, although this is not always the case, particularly for births overseas.

Table 5 Birth registration: completion of variables in birth registration records by country and multiplicity.

Variable	**England** total 10474		**Wales** total 2578		**Scotland** total 2173		**Northern Ireland** total 1615	
	Singletons 10182	multiples 292	singletons 2514	multiples 64	singletons 2116	multiples 57	singletons 1569	multiples 46
Birth weight	10170 (100%)	292 (100%)	2511 (100%)	64 (100%)	Not available		not available	
Mother's place of birth	10182 (100%)	292 (100%)	2513 (100%)	64 (100%)	Not available		not available	
Mother's country of birth code	10181 (100%)[1]	292 (100%)[1]	2512 (100%)[1]	64 (100%)[1]	2116 (100%)	57 (100%)	not available	
Mother's occupation	8043 (79%)	250 (86%)	2010 (80%)	56 (88%)	1742 (82%)	48 (84%)	1170 (75%)	40 (87%)
Mother's occupation code[2]	10181 (100%)	292 (100%)	2514 (100%)	64 (100%)	2116 (100%)	57 (100%)	1569 (100%)	46 (100%)
Mother's occupation code[3]	5976 (59%)	237 (81%)	1473 (59%)	35 (55%)	1692 (80%)	48 (84%)	1532 (97%)	44 (96%)
Mother's employment status	Not available		not available		2116 (100%)	57 (100%)	not available	
Mother's social class	Not available		not available		2116 (100%)	57(100%)	not available	
Mother's nssec	Not available		not available		1692 (80%)	48 (84%)	not available	
Mother's industry	Not available		not available		1450 (69%)	44 (77%)	not available	
Father's place of birth	9410 (92%)	280 (96%)	2239 (89%)	62 (97%)	Not available		not available	
Father's country of birth code	Not available		not available		2116 (100%)	57 (100%)	not available	
Father's occupation	9209 (90%)	272 (93%)	2178 (87%)	62 (97%)	1936 (91%)	57 (100%)	1383 (88%)	42 (91%)
Father's occupation code[2]	10182 (100%)	292 (100%)	2514 (100%)	64 (100%)	2091 (99%)	57 (100%)	1418 (90%)	44 (96%)
Father's occupation code[3]	8726 (86%)	255 (87%)	2092 (83%)	62 (97%)	1874 (89%)	57 (100%)	1418 (90%)	44 (96%)
Father's employment status	Not available		not available		2091 (99%)	57 (100%)	not available	
Father's social class	Not available		not available		2091 (99%)	57 (100%)	not available	
Father's nssec	Not available		not available		2091 (99%)	57 (100%)	not available	
Father's industry	Not available		not available		1571 (74%)	51 (89%)	not available	

[1] has been derived by staff at NPEU from place of birth and ISO3166 coding

[2] includes codes for "other occupations" such as housewives, armed forces, students and codes such as "not classifiable" (e.g. not enough information given, or no father registered, which is not separately recorded in the variables extracted)

[3] "classifiable occupations" only

Hospital Records

Table 6 shows the number of additional variables gained through matching to hospital records. The records from Northern Ireland gained the most general record variables although no 'tail' variables were gained. The number of variables gained in the remaining countries ranged between 16 and 21 for 'tail' variables and between 39 and 55 for general records.

Table 6 Hospital records: number of additional variables gained through matching by country.

Country	Additional variables gained		
	Pregnancy 'tail' Variables	**Baby 'tail' variables**	**'General record' variables**
England	16	5	55
Wales	13	5	52
Scotland	11	5	39
Northern Ireland	0	0	64

Appendix 3, **Table 1** shows the completion of the additional variables gained through matching MCS data to hospital records. These are described briefly below by country.

England

Completeness of the additional 'tail' variables varies from 65% (anaesthetic post delivery) to 99% (number of babies). However, it should be noted that even for variables that are well completed, the proportions with such data reduce markedly when they are calculated out of all MCS babies. For example, 90% of singletons (n=6,113) have antenatal days of stay completed, although only 54% (6,113/11,375) of all MCS singletons have this variable completed i.e. those who consented and had a successful match and had a 'tail' and had this 'tail' variable completed. Completeness for most of the additional 'general record' variables is 100%. Many of the variables for which completeness is less than 100% are those that are not necessarily applicable, such as (additional) diagnosis and operations codes.

Wales

Completeness of the additional 'tail' variables varies markedly from 15% (birth order) to 100% (delivery method and number of previous pregnancies). Again, it should be noted that these proportions will be much lower if calculated for all MCS babies. There were no completed 'tail' variables in the multiples due to none of the multiple matches having 'tails'. Completeness for most 'general record' variables is 100%. Again, many of the variables for which completeness is less than 100% are those that are not necessarily applicable, such as (additional) diagnosis and operations codes.

Scotland

Four of the pregnancy-related variables had a completeness rate of 68% (delivery place change, analgesia given during labour, analgesia given during delivery and most senior doctor present) and the remaining variables had 98-100% completion. All of the baby-related variables had completeness rates above 90%. Completeness for most 'general record' variables is 100%. Again, many of the variables for which completeness is less than 100%

are those that are not necessarily applicable, such as (additional) diagnosis and operations codes.

Northern Ireland
In Northern Ireland, there are no 'tail' variables gained. Completeness for most 'general record' variables is 100%. Again, many of the variables for which completeness is less than 100% are those that are not necessarily applicable, such as (additional) diagnosis and operations codes.

Agreement between MCS and linked records

Methods for assessing agreement between MCS and linked records

Selected variables that were collected both in the MCS and in the birth registration or hospital data were assessed for agreement. We also assessed whether particular factors were associated with agreement, that is, that agreement between records was biased. Where there was less than perfect agreement between these variables, several possible explanations were explored:

i) *That the records being compared do not belong to the same person, that is, that the records are not true matches.*
The probability of false matches was explored in **Appendix 1**. These results suggest that most matches are likely to be true although we cannot rule out some false matches. For the England data, matching was undertaken hierarchically with several different match types, where *match type* 1 was the most stringent and *match type* 5 was the least stringent. We explored whether **match type** was associated with poor agreement. Note that this analysis was done *post hoc*.

ii) *That the data in the hospital records are incorrect.*
Errors in the hospital records may have occurred *randomly*, in which case, agreement would be less than perfect, but no obvious patterns would be present. *Systematic* errors in the hospital records may be detected by exploring whether **hospital** was associated with poor agreement. Other types of systematic error were explored such as the HES data cleaning algorithms (e.g. missing baby's sex in the *tail* record being overwritten with sex from the *general* record for some hospitals).

iii) *That the data in the MCS records are incorrect.*
Errors in the MCS may have occurred *randomly*, in which case, agreement would be less than perfect, but no obvious patterns would be present. *Systematic* errors in the MCS records may be detected by exploring whether variables related to **data collection and entry** (e.g. "wave" or sampling point) were associated with poor agreement. In addition, systematic errors may occur if maternal recall or reporting of the variables is less than perfect. Such recall or reporting bias may be differentially associated with maternal characteristics, and as such, may be detected by exploring whether maternal characteristics were associated with poor agreement.

Agreement between MCS and birth registration birth weight

Birth weight obtained from Birth registration was compared with birth weight as reported for MCS cohort children. Accuracy within 100 g was 92% overall, varying from 94% among British/Irish white mothers to 69-89% for other ethnic groups and was lower among the long-term unemployed and those living in disadvantaged or ethnic wards. [5]

Agreement between MCS and centrally collected hospital records

Selected variables that were collected both in the MCS and in the hospital data were assessed for agreement. Where there was less than perfect agreement between these variables, several possible explanations were explored, as described above. Explanations such as the record being a false match or the field being unreliable need to be considered by users of the additional data gained through matching.

Definitions used in the assessment of agreement

Table 7 shows the definitions of agreement that were used when assessing agreement.

Table 7 Hospital records: definition of agreement between 'tail' variables in MCS and hospital records.

Variable	**Groupings used for assessing agreement**
Gestation at 1st antenatal assessment	Within 2 weeks either side
Delivery method	Normal Forceps Ventouse Assisted breech Elective caesarean Emergency caesarean
Gestational length	Within 2 weeks either side
Number of babies	Exact number
Live or still birth	Live birth Still birth
Sex of baby	Male Female Undetermined
Birth weight	Within 0.2 kg either side
Neonatal care	Yes No

Results on agreement

Tables 8 and 9 show the amount of agreement between the 'tail' variables as reported in the hospital records and the MCS. The interpretation of these results is discussed for each variable in turn in Table 10.

Table 8 Hospital records: agreement between the pregnancy-related 'tail' variables in the hospital records and the MCS, by country and multiplicity

	England Total 6,911		**Wales** Total 2,370		**Scotland** Total 2,004		**Northern Ireland** Total 1,133	
	Singletons 6,797	Multiples 114	Singleton 515	Multiple 0	Singleton 1,978	Multiple 26	Singleton 1,103	Multiple 15
Gestation at 1st antenatal assessment[1]	2152/4721 (46%)	41/78 (53%)	NA	NA	NA	NA	NA	NA
Delivery method	5040/5395 (93%)	66/91 (73%)	364/433 (84%)	NA	1901 (96%)	23 (88%)	NA	NA
Gestational length[1]	5459/5783 (94%)	86/89 (97%)	379/421 (90%)	NA	1938 (98%)	25 (96%)	NA	NA
Number of babies	6113/6732 (91%)	83/99 (84%)	435/435 (100%)	NA	1978 (100%)	26 (100%)	NA	NA

[1] Agreement =equal or plus or minus 2 weeks
NA not applicable

Table 9 Hospital records: agreement between the baby-related 'tail' variables in the hospital records and the MCS, by country and multiplicity

	Singletons	**Multiples**		
		baby 1	**baby 2**	**baby 3**
Country	N (%)	N (%)	N (%)	N (%)
England	**6,797**	**114**	**114**	**3**
Live or still birth	6093/ 6160 (99%)	90/114 (79%)	86/114 (75%)	3 (100%)
Birth weight[2]	5805/ 6174 (94%)	79/90 (88%)	79/86 (92%)	3 (100%)
Sex of baby	5947/6169 (96%)	102/107 (95%)	86/87 (99%)	2/3 (67%)
Wales	**515**	**0**	**0**	**-**
Birth weight	352/434 (81%)	No tail items present		
Sex of baby	435/435 (100%)			
Scotland	**1978**	**26**	**26**	**3**
Live or still birth	1907/1907 (100%)	26 (100%)	26 (100%)	3 (100%)
Birth weight[2]	1931/1978 (97%)	25/26 (96%)	25/26 (96%)	3 (100%)
Sex of baby	1972/1972 (100%)	26 (100%)	25/26 (96%)	3 (100%)
Neonatal care	884/969 (91%)	14/15 (93%)	11/15 (73%)	-

[2] Agreement = equal or plus or minus 0.2k

Analysis guidelines

Hospital records: Reliability of match-types and specific variables

The following points should be considered when analysing the linked data:

i) match type (England only) – obviously, the higher order matches are more likely to be true matches. However, a lower order match does not necessarily mean that the records *did not match* on key variables; in some cases (some of match types 4 and all of match types 6), it was that the matching variables were *missing*.

ii) the distinction between poor data in a *field* and poor data in a *record*. For some variables, it is clear that the data entered are incorrect (e.g. for some fields in some hospitals). However, this does not necessarily mean that the whole matched record is of poor quality. Therefore, one needs to consider each variable separately. The user should look up each variable in the appropriate tables on completeness and agreement.

iii) For some variables, we have created a flag which indicates the records where we think the linked data are of good or bad quality.

Table 10 lists the variables used to assess agreement with recommendations for their use.

Table 10 Hospital records: Recommendations and interpretation of results on agreement.

Variable	Interpretation of results and recommendations
Gestation at 1st antenatal assessment	• Data only available in England and agreement very poor, most probably due to different interpretation of 1st antenatal assessment among women and health professionals, and the fact that in hospital records it was derived from 2 dates, so relies on both of these being accurate. • Recommend that this variable is not used in analysis.
Delivery method	• Data available in all countries in the hospital tail (where tails are completed) and in the operation codes (in the general record, for which completion is extremely high). • In those countries with tails (not Northern Ireland), agreement is at least 84%; note that this is based on perfect agreement between 6 mode of delivery groups. • Further work on this variable (see Quigley et al) shows that for the **operation codes**, completion and agreement are higher: at least 94% using 6 mode of delivery groups and at least 98% using 3 groups (normal, assisted, caesarean).
Gestational length	• Data available in all countries; agreement (within 2 weeks) at least 90% in all countries. • Further data cleaning should be conducted before this variable is used e.g. check for improbable values in hospital records and MCS.
Number of babies	• MCS data assumed to be almost 100% accurate, hence discrepancies assumed to be in hospital records. • Data available in Wales (100% agreement), Scotland (100% agreement) and England. • The singleton records in England in which there was disagreement all came from 19 hospitals (Appendix 4, Table1); the hospital data are probably incorrect. • There is a flag present in the archived data for England that should be used with this variable. It indicates which records come from the 19 hospitals identified above.
Live birth	• MCS data assumed to be almost 100% accurate, hence discrepancies assumed to be in hospital records. • Data only available in Scotland (100% agreement) and England. • The 67 singleton records in England which were classified as stillbirths all came from 5 hospitals (Appendix 4, Table 2); the hospital data are probably incorrect. • There is a flag present in the archived data for England that should be used with this variable. It indicates which records come from the 5 hospitals identified above.
Birth weight	• Data available in all countries except Northern Ireland; agreement (within 0.2 kg either side) at least 81%. • Note that birth weight was used as a matching variable for some of the match types in England and hence for these records, agreement cannot be assessed. • Further data cleaning should be conducted before this variable is used e.g. check for improbable values in hospital records and MCS. • Agreement has also been assessed between MCS and **birth registration** birth weight [5], which found agreement (within 0.1kg either side) for 92% of records although this varied according to ethnicity and ward type.
Baby's sex	• Data available in Wales (100% agreement), Scotland (100% agreement in singletons) and England (many missing values). • There were 222 singleton records in England in which there was disagreement. For 11 of these, the hospital records coded the sex as undetermined; these all occurred in one hospital (11/85 records, 13%; it is implausible to have this proportion undetermined). • The remaining 211 discrepancies were in 44 hospitals. For some hospitals, it was clear that the data were unreliable. For example, all babies were coded as female in two hospitals (possibly as the result of an erroneous HES cleaning rule) with a large number of records (hospital u 150/150; hospital v 59/59) and in three smaller hospitals. In hospital w, 94/123 babies (76%) were coded as male (30/123=24% discrepancies). • After excluding hospitals u, v and w, there were 74 discrepancies. • Some discrepancies due to poor completion in some hospitals, although cannot rule out some discrepancies due to incorrect data in MCS or incorrect matches. • There is a flag present in the archived data for England that should be used with this variable. It indicates which records come from the 3 hospitals identified above.
Neonatal care	• Data only available in Scotland; agreement ranges between 73% and 91%.

Intra-country analysis

Potential caveats of the additional data:

- The additional linked data apply to certain subgroups only:
    - those who gave consent (90% of the MCS mothers)
    - natural mothers (99.7% of the MCS mothers)
    - those delivered in NHS hospitals (estimated as 97.5% of MCS mothers).

  This subgroup is not representative of the MCS and hence, estimates derived using the linked data may be prone to bias. For example, consent was less likely in those living in disadvantaged or ethnic minority wards, those from minority ethnic groups, lone parents and those with higher degrees or no qualifications [6].

- One cannot rule out that a small number of matches are false, in other words, that records for two different people have been linked. This would cause misclassification bias, which would lead to diluted effects (e.g. odds ratios tending towards 1).

- Data in some fields are not well completed and some fields are unreliable. Unreliable fields will not be available for general use by external researchers.

References

1. Shepherd et al. Millennium Cohort Study First Survey: A Guide to the SPSS dataset, May 2003.

2. Plewis I. Millennium Cohort Study First Survey: technical report on sampling, 3rd edition, June 2004.

3. Abrahams C, Davy K. Linking HES maternity records with ONS birth records. *Health Stat Quart* 2002, 13: 22-30.

4. NHS Maternity Statistics, England: 1998-1999 to 2001-2002. 2002. Department of Health, London.

5. Tate AR, Dezateux C, Cole TJ, Davidson L, et al. Factors affecting a mother's recall of her baby's birth weight. *Int J Epidemiol* 2005, 34: 688-95

6. Tate AR, Calderwood L, Dezateux C, Joshi H, et al. Mother's consent to linkage of survey data with her child's birth records in a multi-ethnic national cohort study. *Int J Epidemiol* 2006, 35: 294-8.

Acknowledgements

This work was conducted by:

National Perinatal Epidemiology Unit, University of Oxford
Maria A Quigley, Chris Hockley.

Mailman School of Public Health, Columbia University, New York City
Leslie L. Davidson.

Centre for Longitudinal Studies, Institute of Education, University of London
Gareth Hughes, Lisa Calderwood, Heather Joshi, Rachel Rosenberg

Centre for Paediatric Epidemiology and Biostatistics, Institute of Child Health
Carol Dezateux, Rosemary Tate.

We are grateful to the following people and organisations for their assistance:

ONS: Carole Abrahams, Louise Bailey, Gillian Vaughan, Pauline Wall, Zoe Uren

GRO(S): Ian Brown, Carolyn MacPherson

NISRA: Maire Brolly, Gillian Fegan, Sandy Fitzpatrick

Northgate Solutions: Kate Byram

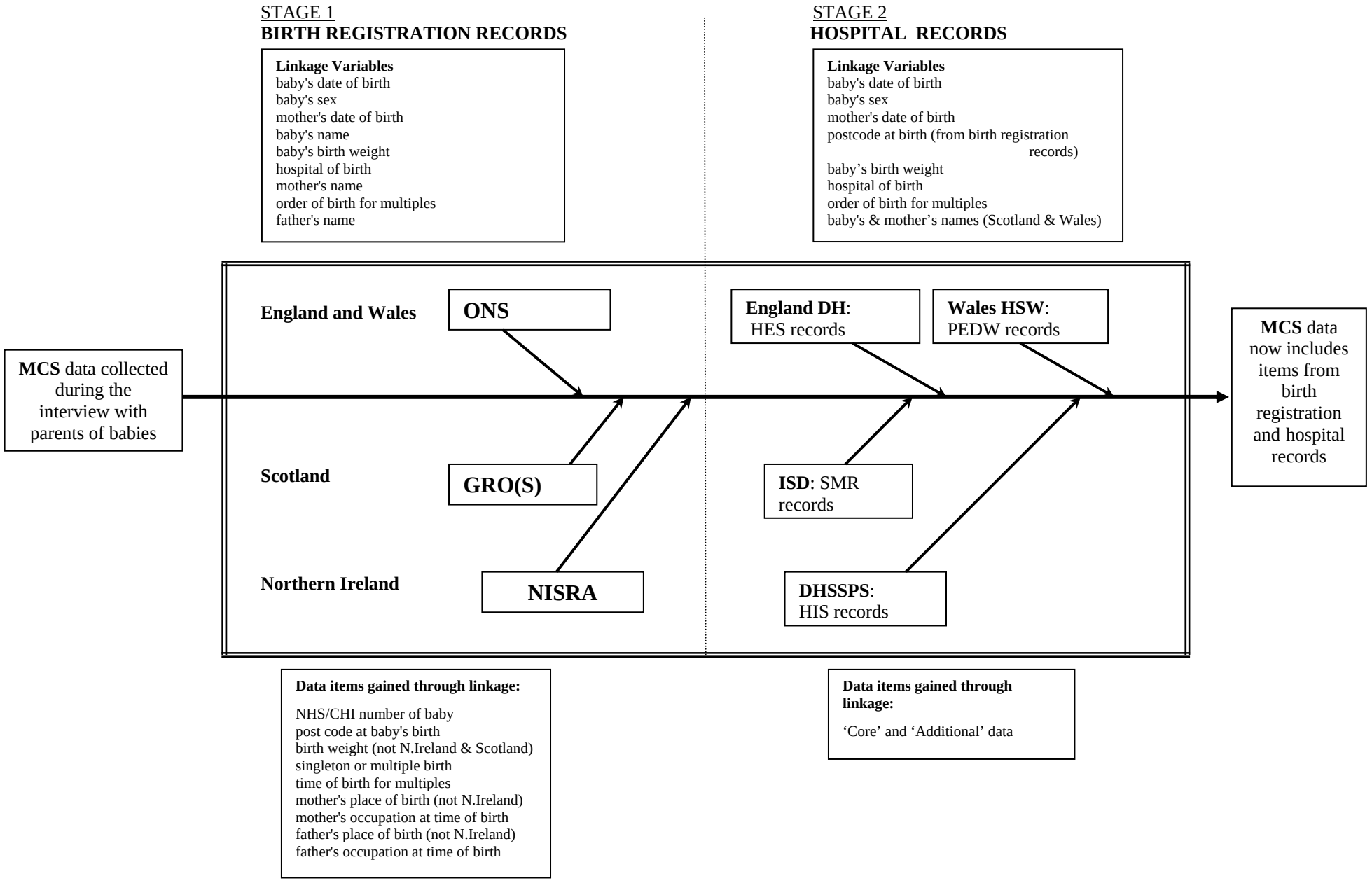
HSW: Gareth John

ISD: Richard Dobbie, Andrew Elders

DHSSPS: Eddie Finn, Sarah Lowe, David Marshall, Jim McColgan, Shaun McCann, Stephen Sharp.

We would also like to thank Yvonne Kelly, Alison Macfarlane and Ziyah Mehta.

Figure 1: Record Linkage for England, Wales, Scotland and Northern Ireland



Appendix 1 - Matching of MCS records with hospital records:

Crude probabilities of matches agreeing by chance for deterministic matching method

These probabilities (expressed as percentages) were estimated using the distribution of the matching variables in records that were known to be non-matches, that is, first in the HES dataset (before any match checking by NPEU) and then in the MCS dataset. By using datasets which included only unique (i.e. non-matching) records, the probability of a match purely by chance was estimated

The probability of two records having the same baby's year and month of birth purely by chance is relatively high due to the study period being about a year (3.8% in HES and 8.3% in MCS). However, the probability of two records having the same whole date of baby's birth is small (0.17% in HES and 0.27% in MCS). For mother's date of birth, the probabilities are much smaller due to the wide variation in maternal age.

The probabilities for the individual matching fields in *Table 1* have been multiplied together in order to obtain probabilities for the different match types (*Table 2*). For example, the probability of two records having the same baby's date of birth and mother's date of birth and baby's sex is <0.0001% in HES and MCS, and would be even lower if the two records had to have the same first three characters of postcode. All of the probabilities in Table 2 are small enough to rule out chance with the exception of match types 4a (0.28% in HES and 0.136% in MCS) and 6 (0.02% in HES and 0.016% in MCS). Once a postcode match is added to match types 4a and 6, we believe that these probabilities would be sufficiently small to make chance unlikely.

However, it should be noted that the estimated probabilities are crude because:

1. We do not have MCS or HES data on postcode and therefore cannot estimate matching probabilities for postcodes or for match types based on postcode.
2. The matching process is complicated by missing values e.g. some records match on certain matching fields but have other matching fields missing.
3. We have estimated the probability for each match type by multiplying the probabilities for the individual matching fields. This is only valid if the individual probabilities are independent.
4. These probabilities give an estimate of the number of false positives (false matches) occurring due to two different individuals having the same matching fields just by chance. False positives may also occur if two different individuals have the same matching fields due to errors in the data. If such errors are random then these probabilities are likely to be small. If such errors are non-random and extensive then these may increase the number of false positives.

Despite these limitations, the crude estimates give an indication as to whether each match type is likely to occur by chance. We think that the probabilities of two records matching purely by chance is extremely small for all match types.

Table 1: Crude probabilities of matches agreeing by chance within HES and MCS for each matching variable.

Matching Variable	HES	MCS
	%	%
Year & Month of baby's DOB	3.800[1]	8.330[1]
Whole baby's DOB	0.172[2]	0.274[2]
Year & Month of mother's DOB	0.280	0.271
Whole mother's DOB	0.020	0.016
Baby's sex	50.000	50.000
Baby's exact birth weight MCS	0.280	0.280
Baby's exact birth weight BR	0.280	0.240

[1] 1/26 for HES (i.e. 26 different month and year combinations) and 1/12 for MCS
[2] 1/582 for HES (i.e. 582 different dates) and 1/365 for MCS

Table 2: Crude probabilities of matches agreeing by chance withinHES and MCS for each match type.

Match type	Matching variables available	The probability that variables will agree purely by chance (%) in HES	in MCS
1	Baby's date of birth, Mother's date of birth Baby's sex *If postcode were included*	 <0.0001 <0.0001	 <0.0001 <0.0001
2	Baby's date of birth, Mother's date of birth Baby's sex Birth weight *If 1st 3 characters of postcode were included*	 <0.0001 <0.0001	 <0.0001 <0.0001
3	a. Baby's date of birth, Baby's sex Year & month of Mother's date of birth *If 1st 3 characters of postcode were included* b. Baby's date of birth, Baby's sex Year & month of Mother's date of birth Birth weight c. Baby's date of birth, Baby's sex Birth weight *If 1st 3 characters of postcode were included*	 0.0002 <0.0002 <0.0001 0.0002 <0.0002	 0.0002 <0.0002 <0.0001 0.0003 <0.0003
4	a. Year & month of Mother's date of birth *If postcode were included* b. Year & month of Mother's date of birth baby's date of birth *If 1st 3 characters of postcode were included*	0.2800 <0.2800 0.0005 <0.0005	0.1362 <0.1362 0.0004 <0.0004
5	a. Mother's date of birth Birth weight b. Mother's date of birth baby's date of birth *If 1st 3 characters of postcode were included*	 0.0001 <0.0001 <0.0001	 <0.0001 <0.0001 <0.0001
6	Mother's date of birth *If postcode were included*	0.0201 <0.0201	0.0162 <0.0162

Appendix 2

Table 1:Hospital records (England): completion of matching variables in matched delivery records for singleton births.

Matching variable	**Match Type**	**Non-missing values No (%)**
Baby's date of birth	1, 2, 3, 4, 5	6,120[1] (72%)
mother's date of birth	1, 2, 3, 4, 5, 6	8,452 (100%)
baby's sex	1, 2, 3	7,155 (85%)
baby's birthweight	2, 3, 5	6,181 (73%)
whole postcode	1, 2, 3, 4, 5, 6	5,685 (67%)
hospital of birth	6	8,414 (99%)
Number of babies		8,458

Table 2: Hospital Records (England): success of matching for singleton and multiple births.

Match type and matching criteria	**No (%)[2] of hierarchical matches**	
	Singletons	**Multiples**
MCS Mothers who gave consent	**10,158**	**260**
1 Baby's DoB, mother's DoB, baby's sex & postcode[1]	3,389 (33%)	160 (62%)
2 Baby's DoB, mother's DoB, baby's sex, first three chars postcode[1] & birthweight	209 (35%)	0 (62%)
3 Baby's DoB, baby's sex & any two of : year & month of mother's DoB, first three chars postcode[1], birthweight.	2,257 (58%)	9 (65%)
4 a. Year & month of mother's DoB & postcode OR b. Year & month of mother's DoB, first three chars postcode[1] & baby's DoB.	1,057 (68%) 207 (70%)	50 (84%)
5 Mother's DoB & either birth weight or baby's DoB & first three chars postcode[1]	124 (71%)	2 (85%)
6 No tail. Mother's DoB, postcode[1] & hospital of birth	1,215 (83%)	10 (89%)
Total matched	8,458 (83%)	231 (89%)

1 Postcode from birth registration matched with postcode from Hospital Episode Statistics
2 Cumulative percentages

Table 3: Hospital records (Wales): completion of matching variables in matched delivery records for singleton births.

Matching variable	Non-missing values No (%)
Baby's date of birth	424 (18%)
mother's date of birth	2,326 (100%)
baby's sex	435 (19%)
baby's birthweight	435 (19%)
whole postcode	2,326 (100%)
hospital of birth	2,326 (100%)
Number of babies	2,326

Table 4: Hospital Records (Wales): success of matching for singleton and multiple births.

Matching criteria	No (%) of matches	
	Singletons	Multiples
MCS Mothers who gave consent	2,545	59
1 Admission date 0-5 days prior to date of baby's birth, mother's date of birth	2,326 (91%)	44 (75%)
Total matched	2,326 (91%)	44 (75%)

Table 5: Hospital Records (Northern Ireland): success of matching for singleton and multiple births.

	No (%)[1] of hierarchical matches	
Matching criteria	**Singletons**	**Multiples**
MCS Mothers who gave consent	**1,711**	**49**
1 Baby's sex, exact birthweight, admission date prior to date of baby's birth & sex & mother's age[2]	1098 (64%)	25 (51%)
2 Baby's sex, exact birthweight, admission date prior to date of baby's birth & sex	1103 (64%)	5 (61%)
Total matched	1103 (64%)	30 (61%)

1 Cumulative percentages
2 Five singleton cases have mother's age 1 year difference

Appendix 3

Table 1: Hospital Records: Completion of variables in the general record and the 'tail' by country.

	England %	**Wales %**	**Scotland %**	**Northern Ireland %**
General record				
Date of admission	100%	100%	100%	100%
Method of admission	100%	100%	100%	100%
Source of admission	100%	100%	100%	100%
Bed days within the data year	100%	NA	NA	100%
Patient classification	100%	100%	100%	100%
Current electoral ward	100%	100%	100%	100%
Diagnosis 1	100%	100%	100%	93%
Diagnosis 2	97%	59%	46%	64%
Diagnosis 3	54%	68%	26%	0%
Diagnosis 4	28%	36%	15%	0%
Diagnosis 5	14%	17%	8%	0%
Diagnosis 6	6%	7%	NA	NA
Diagnosis 7	2%	1%	NA	NA
Diagnosis 8	NA	0%	NA	NA
Diagnosis 9	NA	0%	NA	NA
Date of discharge	100%	100%	96%	100%
Destination on discharge	100%	100%	100%	100%
Method of discharge	100%	100%	100%	100%
Patient's date of birth	100%	100%	NA	5%
Age at start of episode	100%	NA	NA	100%
Age at end of episode	100%	100%	NA	NA
Episode duration	100%	100%	NA	100%
Date episode ended	100%	100%	100%	100%
Episode start date	100%	100%	100%	100%
Ethnic group	70%	100%	NA	NA
Intended management	100%	100%	100%	0%
Main specialty	100%	100%	100%	100%
Date of operation 1	96%	97%	28%	93%
Date of operation 2	63%	70%	8%	80%
Date of operation 3	32%	34%	2%	46%
Date of operation 4	14%	16%	0%	24%
Date of operation 5	NA	3%	NA	NA
Date of operation 6	NA	1%	NA	NA
Date of operation 7	NA	0%	NA	NA
Operation code 1	98%	97%	28%	93%
Operation code 2	66%	70%	1%	80%
Operation code 3	33%	34%	8%	46%
Operation code 4	14%	16%	9%	24%
Operation code 5	NA	3%	2%	NA
Operation code 6	NA	1%	0%	NA
Operation code 7	NA	0%	0%	NA
Operation code 8	NA	NA	0%	NA
Post operation duration (days)	96%	NA	NA	NA
Pre operation duration	96%	NA	NA	100%

	England %	**Wales %**	**Scotland %**	**Northern Ireland %**
Provider code	100%	100%	100%	100%
Purchaser code	100%	100%	100%	NA
County of residence	100%	NA	NA	100%
Government Office Region	100%	NA	NA	NA
Health Authority of residence	100%	NA	NA	NA
Local authority district	100%	NA	NA	NA
Regional Off of residence	100%	NA	NA	98%
Region of treatment(RO)	100%	NA	NA	100%
Sex	100%	100%	NA	100%
Beginning of spell	100%	100%	NA	NA
Spell duration	98%	100%	NA	NA
End of spell	100%	100%	NA	NA
Treatment specialty	94%	100%	NA	100%
Ward type at start of episode	99%	NA	NA	NA
District of treatment	100%	NA	NA	NA
Cost per day	100%	NA	NA	NA
HRG recent value	100%	NA	NA	93%
HRG original value	100%	NA	NA	93%
Notional total cost	100%	NA	NA	NA
Cost of treatment	100%	NA	NA	NA
Significant facility	NA	NA	100%	NA
Clinical facility at start	NA	NA	6%	NA
Marital Status	NA	NA	NA	100%
'tail' - pregnancy related variables				
First antenatal assessment date	87%	75%	98%	NA
Antenatal days of stay	90%	90%	NA	NA
Postnatal days of stay	89%	89%	NA	NA
Gestation at first antenatal assessment	75%	NA	NA	NA
Delivery place type (intended)	85%	84%	100%	NA
Delivery place type (actual)	87%	84%	68%	NA
Reason for change in delivery place	76%	41%	NA	NA
Delivery onset method	94%	78%		NA
Delivery method	87%	100%	100%	NA
Duration of labour	NA	NA	99%	NA
Analgesia given during labour	NA	NA	68%	NA
Anaesthetic given during delivery	85%	43%	-	NA
Analgesia given during delivery	NA	NA	68%	NA
Anaesthetic given post delivery	65%	32%	NA	NA
Status of person conducting delivery	90%	84%	68%	NA
Gestation length	86%	84%	100%	NA
Number of previous pregnancies	80%	100%	100%	NA
Number of babies	99%	84%	100%	NA
Mother's age at birth	91%	NA	NA	NA
'tail' - baby related variables				
Resuscitation method	86%	71%	91%	NA
Birth order	94%	70%	NA	NA
Live or still birth	91%	84%	96%	NA
Birth weight	91%	84%	100%	NA
Sex of baby	91%	84%	100%	NA
Level of neonatal care	NA	NA	100%	NA

Appendix 4

Table 1: Hospital records (England): patterns of disagreement in those hospitals where MCS and hospital records disagreed for "no. babies".

Hospital	No. records in hospital [non-missing]	Number of babies					
		Agreement	Disagreement				
		1 No. (%)	2 No. (%)	3 No. (%)	4 No. (%)	5 No. (%)	6 No. (%)
A	19 [19]	1 (5%)	0	0	0	0	18 (95%)
B	5 [5]	4 (80%)	1 (20%)	0	0	0	0
C	23 [5]	9 (39%)	1 (4%)	0	0	0	13 (57%)
D	189 [189]	118 (62%)	0	0	0	71 (38%)	0
E	29 [27]	0	0	0	0	0	27 (100%)
F	3 [3]	2 (66%)	0	0	0	0	1 (33%)
G	29 [29]	0	29 (100%)	0	0	0	0
H	37 [37]	34 (87%)	0	0	0	0	3 (13%)
I	106 [105]	103 (98%)	0	0	0	0	2 (2%)
J	5 [5]	4 (80%)	0	0	0	0	1 (20%)
K	54 [53]	0	0	0	0	0	53 (100%)
L	125 [125]	4 (3%)	0	54 (43%)	0	0	67 (54%)
M	49 [49]	0	0	0	0	0	49 (100%)
N	8 [7]	0	0	0	0	0	7 (100%)
O	27 [26]	17 (65%)	2	0	0	0	7 (27%)
P	71 [71]	0	0	0	0	0	71 (100%)
Q	60 [58]	44 (76%)	0	0	0	0	14 (24%)
R	44 [44]	14 (32%)	0	0	30 (68%)	0	0
S	176 [176]	98 (56%)	0	0	0	0	78 (44%)

Table 2: Hospital records (England): patterns of disagreement in those hospitals where MCS and hospital records disagreed for "live-birth".

Hospital	No. records in hospital	Agreement (HES says live birth) No. (%)	Disagreement (HES says stillbirth) No. (%)	HES missing No. (%)
1	17	0 (0%)	2 (12%)	15 (88%)
2	39	0 (0%)	18 (46%)	21 (54%)
3	125	88 (70%)	30 (24%)	7 (6%)
4	12	8 (67%)	4 (33%)	0 (0%)
5	41	0 (0%)	13 (32%)	28 (68%)