Routinely collected English birth datasets: comparisons and recommendations for reproductive epidemiology

Rebecca E. Ghosh,¹ Danielle C. Ashworth¹, Anna L. Hansell^{1,2}, Kevin Garwood¹, Paul Elliott,^{1,2} Mireille B. Toledano¹

¹UK Small Area Health Statistics Unit, MRC-PHE Centre for Environment and Health, School of Public Health, Imperial College London, W2 1PG, UK

²Imperial College Healthcare NHS Trust, London, UK

Corresponding author – Mireille B. Toledano

MRC-PHE Centre for Environment and Health

School of Public Health

Imperial College London

W2 1PG

UK

Email: m.toledano@imperial.ac.uk

Tel: 020 7594 3298

Keywords: Routine birth data, reproductive epidemiology, hospital episode statistics, office for national statistics, data quality

Word count: 2875 words

ABSTRACT (247 words)

Background:

In England there are four national routinely collected datasets on births: Office for National Statistics (ONS) births based on birth registrations; Hospital Episode Statistics (HES) deliveries (mothers information); HES births (babies information); and NHS Numbers for Babies (NN4B) based on ONS births plus gestational age and ethnicity information. This study describes and compares these data, with the aim of recommending the most appropriate dataset(s) for use in epidemiological research and surveillance.

Methods:

We assessed the completeness and quality of the datasets in relation to use in epidemiological research and surveillance and produced detailed descriptive statistics on common reproductive outcomes for each dataset including temporal and spatial trends.

Results:

ONS births is a high quality complete dataset but lacks interpretive and clinical information. HES deliveries showed good agreement with ONS births but HES births showed larger amounts of missing or unavailable data. Both HES datasets had improved quality from 2003 onwards, but showed some local spatial variability. NN4B showed excellent agreement with both ONS and HES deliveries for the years available (2006-2010). Annual number of births increased by 17.6% comparing 2002 to 2010 (ONS births). Approximately 6% of births were of low birth weight (2.6% term low birth weight) and 0.5% were stillbirths

Conclusions:

Routinely collected data on births provide a valuable resource for researchers. ONS and NN4B offer the most complete and accurate record of births. Where more detailed clinical information is required, HES deliveries offers a high quality dataset that captures the majority of English births.

What is already known on this subject?
<ul style="list-style-type: none">• Routinely collected birth datasets provide an important resource for epidemiological studies and for surveillance of reproductive health.• Of the four national sources of birth data ONS births and NN4B offer the most complete and accurate record of all births in England but do not provide the detailed clinical information held in HES.
What this study adds?
<ul style="list-style-type: none">• HES deliveries are recommended for use over HES births but researchers should undertake a descriptive analysis of the data to identify any temporal or spatial trends.• Policies are required to reduce the high burden of permissions and information governance required to obtain and link birth datasets.

INTRODUCTION

Routinely collected data on births are a valuable resource for use in epidemiological studies of reproductive outcomes and surveillance [1,2]. In England there are four national births datasets: Office for National Statistics (ONS) births based on birth registrations; Hospital Episode Statistics (HES) deliveries; HES births; and NHS Numbers for Babies (NN4B) which is based on ONS births plus gestational age and ethnicity information.

ONS births consists of information on births in England and Wales (live or still) registered within 42 days of birth, a statutory requirement. ONS births is a complete, high quality dataset that holds some socio-demographic information, but lacks key information such as gestational age. ONS provide detailed metadata as well as producing annual publications on a range of summary birth statistics and trends [3]. This dataset has also been used in studies of reproductive health and environmental exposures [1,2,4–6], socio-demographic effects [7–9], temporal trends in birth weight [10,11] and survival [12].

HES from the Health and Social Care Information Centre (HSCIS), documents all admissions into English NHS hospitals and facilities funded by the NHS and routinely publishes descriptive statistics and data quality summaries of their maternity data [13]. The HES maternity records are a subset of HES and include two datasets: HES births (babies record), and HES deliveries relating to the birth process (mothers record with information on each baby). HES deliveries contains detailed clinical information and has been used to investigate obstetric surgery outcomes and practice [14–17]. HES births has been mainly used for methodological papers creating linked birth cohorts [18] and linkage with other routine datasets [19,20].

NN4B was created to allocate NHS numbers to babies who are notified (within a few days of birth) to a Central Issuing System and for 2006 onwards NN4B is available from ONS. NN4B is a high quality record of births with additional key variables including gestational age [21] and has been linked to birth and death registrations by ONS to produce gestation-specific infant mortality statistics [22]. From 2015 the NN4B functions have been replaced by the Personal Demographics Service (PDS) on the NHS Spine but the information currently being collected will remain comparable to NN4B.

The specific aims of this study were to: 1) assess, for the first time, the quality of all four national data sources on birth outcomes; 2) to produce comparative statistics for each dataset for several common outcomes in reproductive epidemiology; and 3) make recommendations to researchers on the most appropriate dataset(s) for use in epidemiological studies.

METHODS

Data were extracted from ONS births registrations and HES maternity records for all English births for calendar years 2002 to 2010; (NN4B from 2006 the earliest available year from ONS). For ONS births and NN4B, each record relates to one birth. In HES maternity, records relate to an episode of care during pregnancy rather than a birth and contain variables that are also held in standard HES records. HES deliveries and HES births hold space for up to nine additional fields known as a “baby tail” in which variables relate to the delivery and babies.

Supplementary Table 1 and Figure 1 document an exploration of the different criteria, filtering conditions and de-duplication that can be used to define a birth within HES. This was conducted in accordance with previously published papers [18,19], reports [23] and personal communication with the HES team.

Data analysis

Variables available, total numbers of births and missing data in each dataset were compared for the whole period (2002-2010) and by year. Descriptive statistics were produced for the following four common adverse birth outcomes [24,25]:

1. Low birth weight (LBW): Live singleton births with a birth weight between 200 and 2500 grams
2. Stillbirths: Births coded as stillbirths occurring at or greater than 24 weeks of pregnancy
3. Term LBW: Live singleton births with a birth weight between 200 and 2500 grams born at or more than 37 weeks of pregnancy
4. Pre-term delivery (PTD): Live singleton birth occurring before 37 weeks of pregnancy but greater than 10 weeks of gestation

Variability in rates and counts was investigated by maternal age, plurality, region, deprivation (quintiles of the Carstairs index [26] of Census Output Areas (COA) 2001 defined by residential postcode), birth weight, sex, parity (number of previous children), ethnicity, previous pregnancies, delivery method, delivery place and gestational age.

TO BE INCLUDED AS A TEXT BOX *Geography: Census Output Areas (COA) are the smallest geographical areas for which census data are published. They are created from Census data using clusters of adjacent postcodes and are designed to have similar characteristics and population sizes (on average around 100 people). Lower layer super output areas (LSOA) are geographical areas built up from groups of adjacent COAs with similar characteristics and for the 2001 census LSOAs had a mean population of 1,514 [27].*

For the two most complete datasets (ONS and HES deliveries) the numbers of live and stillbirths and the average birth weight were calculated at a small area level (Lower layer Super Output Area 2001 – LSOA) across: a) England and b) two regions, the North-East, and Greater London. The rates of live births per 10,000 population (using ONS mid-year population estimates) of the two regions were then mapped at LSOA level.

All data handling and analysis was performed in R version 2.14.2 and STATA version 13 (Stata Corp., College Station, Texas, US); maps were produced using Arc GIS 10.1 (Environmental Systems Research Institute, CA).

RESULTS

All datasets contain information on birth status and weight, sex, mothers' date of birth, and residential postcode (Table 1). ONS and NN4B have residential address while the HES maternity datasets only have information at postcode level. The HES datasets and NN4B include information on gestational age and ethnicity, with HES providing the mothers ethnic group and NN4B providing the ethnic category of the baby as defined by the mother. HES maternity also provides additional clinical information.

Total births and time trends

There were clear differences in capture across the study period (Supplemental figure 2) and in 2001 HES deliveries captured only 73.0% of ONS. Capture was much higher thereafter, so 2001 was excluded from subsequent comparisons. From 2002 to 2010, ONS recorded 5,727,407 births, HES deliveries 5,545,905 and HES births 5,534,194, while NN4B recorded 3,333,154 in 2006-2010 (Table 2). From 2002 to 2010, HES deliveries captured 96.8% of all ONS births, HES births captured 96.6% and for 2006-10 NN4B captured 99.8%.

ONS had few missing or unavailable data except for parity which was only recorded for married mothers (51.6% of births) and gestational age (stillbirths only); NN4B also had few missing data (Table 2). The HES datasets had more variables with larger proportions of missing data than ONS, with HES births being worse than HES deliveries. ONS and NN4B had little variation in missing data

by year, both HES datasets showed a decrease in missing data over time but with a spike in missing data in 2007 (Supplementary Table 2). For HES births the sex of the baby was not collected from 2003 and for key variables there was an increase in missing data over time.

A comparison of selected variables for each dataset is presented in Supplemental Table 3. For total births the HES datasets were broadly consistent with ONS and NN4B data, with fewer multiple births (HES deliveries 2%; ONS 3.1%) and more female births (HES deliveries 49.6%; ONS 48.7%).

Low birth weight (LBW) births and stillbirths

Similar proportions of births were recorded as LBW in all datasets (ONS =5.9%; HES deliveries=6%; HES births 6.3%; NN4B=5.5%) (Supplemental Table 4) and all showed a decreasing trend in LBW rates (Figure 1A). When comparing specific variables in each dataset for LBW births only, HES deliveries was similar to ONS but HES births was not, especially when comparing sex and region. When comparing the characteristics of LBW births with those of all births (Supplemental table 3), LBW births were more likely to be Asian (NN4B Asian LBW births =16.9%; NN4B Asian births =10.4%), delivered more frequently by Caesarean section and in the most deprived Carstairs quintile.

Similar proportions of all births were stillborn in all datasets (ONS 0.5%; HES deliveries 0.6%; HES births 0.5%; NN4B 0.5%) (Supplemental Table 4). Between 2002-2010 there was a slight decrease in the annual rate of stillbirths with a large peak in 2007 in HES (Figure 1B). This was due to reporting issues with one NHS trust (which has previously been reported in HES data quality notes) and after excluding this trust, the rates of stillbirths in HES were lower than in ONS. When comparing specific variables in each dataset for stillbirths only, HES deliveries was generally similar to ONS, but HES births showed larger discrepancies. Compared to all births (Supplemental Table 3) stillbirths were

more often multiple births (ONS multiple births=3.1%, ONS multiple stillbirths=7.7%), in deprived areas, of LBW and from non-white ethnic groups.

Term low birth weight and pre-term deliveries (PTD)

Analysis of term LBW and PTD requires information on gestational age which is held in the HES datasets and NN4B only (Supplemental Table 5). Using data for 2006-10, the proportion of live singleton births that were term LBW varied from 2.5% in NN4B to 2.7% in HES births. Compared to all births (Supplemental table 3) term LBW babies were more likely to be female, Asian (HES deliveries Asian births = 11.5%, HES deliveries Asian term LBW births = 23.1%), and from deprived areas.

The two HES datasets recorded a similar, higher proportion of PTD than NN4B (HES deliveries=7.4%; NN4B=5.9%) (Supplemental table 5) and the HES deliveries dataset was more similar to NN4B than HES births. Compared with all births, PTD were more likely to be LBW (HES deliveries LBW births=5.8%; HES deliveries pre-term LBW births=45.9%) and from the most deprived areas.

Regional spatial analysis:

The national HES deliveries dataset had fewer live births per LSOA than ONS (HES mean 122.8 live births per LSOA; ONS=175.1) with this difference being most marked in London (Table 3). Stillbirth counts at LSOA level were similar in both datasets at a national level (mean 0.9 stillbirths per LSOA), but lower in HES deliveries in the North East and London. The birth weight data at LSOA level showed good agreement between HES deliveries and ONS but differences were observed in the variability. A similar pattern was seen for the North East and London regions.

The spatial distribution of the rates of live births in London and the North East at LSOA level (2002-2010) is shown in Figure 2. In London there were clear spatial differences in the live birth rates, with

the South East and North West of London showing particularly low rates of HES deliveries compared with ONS births (Figure 2A). In the North East the distribution of births by LSOA was broadly similar in both datasets (Figure 2B).

Discussion:

This is the first study to provide a detailed assessment of the quality of reproductive health data from all four national routine births datasets in England. Overall the ONS births dataset is the most complete and accurate record of all births in England (2002-2010) and NN4B is a valuable enhancement to this dataset. HES deliveries is more complete than HES births and captures the majority of English births (96.8%) with good comparability to ONS but still has inaccuracies relating to missing data resulting in temporal and spatial anomalies. However HES deliveries offers detailed clinical information that cannot be obtained from the ONS datasets.

Descriptive statistics and trends for the birth outcomes were broadly similar for ONS, NN4B and HES deliveries, but less so for HES births due to missing data. The prevalence of LBW babies in ONS (5.9%) was similar to a WHO estimate for similar European countries (6.6%) [28], as was the prevalence of stillbirths in ONS (0.5%) which also was consistent with other European countries (<1%)[29]. The prevalence of PTD in the NN4B data (5.9%) similar to recent 2010 estimates for other Northern European Countries (5%)[30]. Known risk factors for LBW, preterm delivery and stillbirths include deprivation and non-white ethnicity [31]; this was consistent with our results which found term LBW, LBW births, PTD and stillbirths were more likely in non-white ethnic groups and in the most deprived Carstairs quintiles.

Our recommendations for those considering using the four national routine births datasets in England for epidemiological studies of birth outcomes are:

- For studies where clinical and lifestyle data are not required e.g. birth rates/prevalence studies, time trend studies etc., then ONS birth registrations is preferred.
- For studies that require information on gestational age and potential confounders such as ethnicity, NN4B is preferred but currently only available from ONS from 2006 onwards.
- If clinical or pre-2006 information is needed HES deliveries is preferred over HES births unless, for example, information on the child's ethnicity is required over mother's ethnicity.
- Temporal and spatial trends in the HES data should be thoroughly explored before use especially if HES data prior to 2002 are to be used.
- Any spatial and temporal trends identified should be interpreted in the light of changes in reporting.
- Despite information governance and technical challenges, linkage between datasets has the greatest potential to provide the richest and best quality datasets for use in research.

Previous studies of birth outcomes in England have primarily used ONS [1,2,7–12], and HES data have seldom been used for peer reviewed research papers [14–17,32]. It is unclear why the HES datasets have been under-utilised but could be due to concerns over data quality or being a more complicated dataset to work with. Differences in how the data are collected may also influence the choice of dataset.

While ONS birth registrations have remained consistently high quality, the HES dataset had poorer capture and more missing data in earlier years, particularly pre-2002. Completeness has improved considerably and HES currently captures almost all English births in hospitals, although it does not record births outside NHS hospitals (e.g. 2.8% of births occurring at home [33] or in private hospitals). Moreover, geographical identifiers in all the datasets are based on residential postcode of the mother. ONS and NN4B capture English resident mothers who give birth in Welsh, Scottish or

Northern Irish hospitals. HES will not capture English mothers who give birth outside England, although numbers of these are likely to be small.

HES datasets remain susceptible to data artefacts due to the nature of their collection and recording processes across many different hospitals. HES data quality is investigated and reported in the HES data quality notes. There are various methods for selecting and de-duplicating HES maternity records and the choice of method may influence the final dataset. The Dr Foster method used in this paper [34] is not the only available method[35]. The apparent large peak in stillbirths in 2007 was the result of one NHS trust recording 99% of all its delivery episodes as stillborn and was reported in the annual HES data quality note [36]. The increase in missing HES data for 2007 is related to cessation of intensive manual data cleaning for 2007-08 [36].

While the HES deliveries data on a national scale were similar to ONS, we found spatial variations at small area level. Low rates were observed in the South East of London caused by the under or lack of reporting of births by several hospitals. One method to deal with any variations in quality in HES data is to focus research or surveillance only on hospitals with high completeness of recording [18]; another is to link birth datasets. Linking ONS to HES deliveries would combine the completeness of ONS with additional information from HES. Pilot studies testing linkage between HES and ONS records have found that a high rate of linkage can be achieved [37]. However the linkage rate will depend on the years of data investigated, with the most recent pilot studies (2005-2007) able to link between 91-93% of HES deliveries to ONS births [19,20]. ONS routinely link infant mortality records with births to produce statistics on infant and perinatal mortality [38] as well as linking NN4B with births to produce gestation-specific infant mortality statistics [22].

While administrative datasets are a rich data source for epidemiological studies, gaining access can be a slow process taking many months. Access to routinely available births datasets that are not

publicly available (i.e. with sensitive and/or personal information) is only possible with appropriate ethical approval, Health Research Authority Confidentiality Advisory Group governance approval, and data provider approval in place. Researchers will also need to use approved suitably secure facilities, either at their own institution or those provided by ONS or the Administrative Data Service. Changes to legislation and/or data provider changes may introduce further delays to obtaining data; HSCIC updating of data access processes as part of the Health and Social Care act 2012 and issues relating to the introduction of the care.data project have recently resulted in substantial delays. Linkage between HES and ONS datasets is not available routinely and there are additional technical challenges related to record matching and validation. Due to these constraints it is currently more common for researchers to use only one birth dataset, therefore reducing the possible data coverage or depth of clinical information.

CONCLUSION

Routine birth datasets in England provide a valuable resource for epidemiological research on birth outcomes, surveillance of reproductive trends and provision of maternity services. The NN4B dataset appears to be a promising addition for years from 2006, as it has the quality and coverage of ONS births but includes gestational age and ethnicity. The HES deliveries dataset, currently under-used, contains rich clinical information unavailable elsewhere but an appreciation of potential data anomalies is important for researchers. Streamlining data access procedures and routine linkage between these datasets would provide the best use of resources possible and improve use of these data by the research community.

Funding:

The work of the UK Small Area Health Statistics Unit is funded by Public Health England as part of the MRC-PHE Centre for Environment and Health, funded also by the UK Medical Research Council.

Grant number: MR/L01341X/1

Dr. Elliott is an NIHR Senior Investigator and is supported by the MRC-PHE Centre for Environment and Health, the Imperial College Healthcare NHS Trust, the NIHR Imperial College Biomedical Research Centre, and the NIHR Health Protection Research Unit on Health Impact of Environmental Hazards.

Ethics and Information Governance statement:

SAHSU holds approvals from the National Research Ethics Service - reference 12/LO/0566 and 12/LO/0567 - and from the Health Research Authority Confidentiality Advisory Group (HRA-CAG) for Section 251 support (HRA - 14/CAG/1039).

Contributors:

MBT conceived the study. KG was responsible for data extraction and KG and DA were responsible for its cleaning and preparation. DA and REG carried out the statistical analyses and drafted the initial report. REG carried out the mapping. The analyses were interpreted by REG, DA, AH, PE and MBT. All co-authors revised the report and approved the final version. MBT is the guarantor of this paper.

Acknowledgements:

Hospital Episode Statistics data are copyright © 2014, re-used with the permission of the Health and Social Care Information Centre. All rights reserved.

The ONS births data used were supplied by the Office for National Statistics (ONS), derived from the national birth registrations.

Competing Interests:

Dr. Ghosh, Dr. Garwood, Dr. Ashworth and Dr. Toledano have nothing to declare.

Dr. Elliott and Dr. Hansell report grants from the Medical Research Council & Public Health England, during the conduct of the study.

References:

- 1 Toledano MB, Nieuwenhuijsen MJ, Best N, *et al.* Relation of trihalomethane concentrations in public water supplies to stillbirth and birth weight in three water regions in England. *Environ Health Perspect* 2005;**113**:225–32. doi:10.1289/ehp.7111
- 2 Elliott P, Briggs D, Morris S, *et al.* Risk of adverse birth outcomes in populations living near landfill sites. *BMJ* 2001;**323**:363–8. <http://www.ncbi.nlm.nih.gov/pubmed/14702842>
- 3 Office for National Statistics. Birth summary tables, England and Wales. 2015. <http://www.ons.gov.uk/ons/rel/vsob1/birth-summary-tables--england-and-wales/index.html>
- 4 Iszatt N, Nieuwenhuijsen MJ, Bennett JE, *et al.* Trihalomethanes in public drinking water and stillbirth and low birth weight rates: an intervention study. *Environ Int* 2014;**73C**:434–9. doi:10.1016/j.envint.2014.08.006
- 5 Nieuwenhuijsen MJ, Toledano M, Bennett J, *et al.* Chlorination Disinfection By-Products and Risk of Congenital Anomalies in England and Wales. *Environ Health Perspect* 2008;**116**:216–22.
- 6 Elliott P, Toledano MB, Bennett J, *et al.* Mobile phone base stations and early childhood cancers: case-control study. *Br Med J* 2010;**340**:c3077–c3077. doi:10.1136/bmj.c3077
- 7 Messer J. An analysis of the socio-demographic characteristics of sole registered births and infant deaths. *Heal Stat Q* 2011;**79**:107. doi:10.1057/hsq.2011.9
- 8 Moser K, Li L, Power C. Social Inequalities in Low Birth Weight in England and Wales: Trends and Implications for Future Population Health. *J Epidemiol Community Heal* 2003;**57**:687–91. <http://www.jstor.org/stable/25570134> (accessed 9 Dec2014).
- 9 Dibben C, Sigala M, Macfarlane A. Area deprivation, individual factors and low birth weight in England: is there evidence of an ‘area effect’? *J Epidemiol Community Health* 2006;**60**:1053–9. doi:10.1136/jech.2005.042853
- 10 Bakeo A. Trends in live births by mother’s country of birth and other factors affecting low birthweight in England and Wales, 1983–2001. *Heal Stat Q* Published Online First: 2004. file:///C:/Users/rghosh/Desktop/hsq2_tcm77-141910.pdf (accessed 9 Dec2014).
- 11 Alberman E. Are our babies becoming bigger? *J R Soc Med* 1991;**84**:257–60. <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1293219&tool=pmcentrez&rendertype=abstract> (accessed 9 Dec2014).
- 12 Alberman E, Botting B. Trends in prevalence and survival of very low birthweight infants, England and Wales: 1983–7. *Arch Dis Child* 1991;**66**:1304–8. doi:10.1136/adc.66.11.1304
- 13 The Health and Social Care Information Centre. NHS Maternity Statistics - England, 2013–14. 2015. <http://www.hscic.gov.uk/searchcatalogue?productid=17198&q=%22nhs+maternity+statistics%22&topics=0%2fHospital+care&sort=Relevance&size=10&page=1#top>
- 14 Bragg F, Cromwell DA, Edozien LC, *et al.* Variation in rates of caesarean section among English

- NHS trusts after accounting for maternal and clinical risk: cross sectional study. *Br Med J (Clin Res Ed)* 2010;**341**:c5065. doi:10.1136/bmj.c5065
- 15 Gurol-Urganci I, Cromwell DA, Edozien LC, *et al.* The timing of elective caesarean delivery between 2000 and 2009 in England. *BMC Pregnancy Childbirth* 2011;**11**:43. doi:10.1186/1471-2393-11-43
 - 16 Onwere C, Gurol-Urganci I, Cromwell D a., *et al.* Maternal morbidity associated with placenta praevia among women who had elective caesarean section. *Eur J Obstet Gynecol Reprod Biol* 2011;**159**:62–6. doi:10.1016/j.ejogrb.2011.07.008
 - 17 Middle C, MacFarlane A. Labour and delivery of ‘normal’ primiparous women: analysis of routinely collected data. *Br J Obstet Gynaecol* 1995;**102**:970–7. doi:10.1111/j.1471-0528.1995.tb10904.x
 - 18 Murray J, Saxena S, Modi N, *et al.* Quality of routine hospital birth records and the feasibility of their use for creating birth cohorts. *J Public Health (Bangkok)* 2013;**35**:298–307. doi:10.1093/pubmed/fds077
 - 19 Dattani N, Datta-Nemdharry P, Macfarlane A. Linking maternity data for England, 2005-06: methods and data quality. *Heal Stat Q* 2011;**53**–79. doi:10.1057/hsq.2011.3
 - 20 Dattani N, Datta-Nemdharry P, Macfarlane A. Linking maternity data for England 2007: methods and data quality. *Heal Stat Q* 2012;**4**–21.
 - 21 Moser K, Hilder L. Assessing quality of NHS Numbers for Babies data and providing gestational age statistics. *Heal Stat Q* 2008;**15**–23. <http://www.ncbi.nlm.nih.gov/pubmed/18351024>
 - 22 Office for National Statistics. Gestation-specific infant mortality in England and Wales. 2014. <http://www.ons.gov.uk/ons/rel/child-health/gestation-specific-infant-mortality-in-england-and-wales/index.html>
 - 23 Care Quality Commission. Care Quality Commission Periodic Review - Maternity Data Quality Indicator specifications. 2010.
 - 24 Savitz D a., Hertz-Picciotto I, Poole C, *et al.* Epidemiologic measures of the course and outcome of pregnancy. *Epidemiol Rev* 2002;**24**:91–101. doi:10.1093/epirev/mxf006
 - 25 Kramer MS. The epidemiology of adverse pregnancy outcomes: an overview. *J Nutr* 2003;**133**:1592S – 1596S.
 - 26 Carstairs V, Morris R. Deprivation: explaining differences in mortality between Scotland and England and Wales. *BMJ* 1989;**299**:886–9. <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1837760&tool=pmcentrez&rendertype=abstract> (accessed 18 Jul2014).
 - 27 Office for National Statistics. Changes to Output Areas and Super Output Areas in England and Wales, 2001 to 2011. 2012;**1**–13.
 - 28 United Nations Children’s Fund and World Health Organisation. Low Birthweight: Country, regional and global estimates. 2004.
 - 29 Say L, Donner A, Gulmezoglu AM, *et al.* The prevalence of stillbirths: a systematic review. *Reprod Health* 2006;**3**:1. <http://www.reproductive-health-journal.com/content/3/1/1>
 - 30 Blencowe H, Cousens S, Oestergaard MZ, *et al.* National, regional, and worldwide estimates of preterm birth rates in the year 2010 with time trends since 1990 for selected countries: A systematic analysis and implications. *Lancet* 2012;**379**:2162–72. doi:10.1016/S0140-6736(12)60820-4
 - 31 Taylor-Robinson D, Agarwal U, Diggle PJ, *et al.* Quantifying the impact of deprivation on preterm births: A retrospective cohort study. *PLoS One* 2011;**6**:2–9. doi:10.1371/journal.pone.0023163
 - 32 O’Sullivan J, Iyer S, Taylor N, *et al.* Congenital adrenal hyperplasia due to 21-hydroxylase deficiency is associated with a prolonged gestational age. *Arch Dis Child* 2007;**92**:690–2. doi:10.1136/adc.2006.112029
 - 33 Birthplace in England Collaborative Group. Perinatal and maternal outcomes by planned

- place of birth for healthy women with low risk pregnancies: the Birthplace in England national prospective cohort study. *Br Med J* 2011;**343**:d7400–d7400. doi:10.1136/bmj.d7400
- 34 Aylin P, Bottle A, Jen M. HSMR mortality indicators. 2010.
[http://www.nhs.uk/scorecard/Documents/HSMR methodology 09 November.pdf](http://www.nhs.uk/scorecard/Documents/HSMR%20methodology%2009%20November.pdf)
- 35 The Health and Social Care Information Centre, Clough J. Secondary Uses Service (SUS)
 Guidance on Data Duplication Version. 2013. [http://www.hscic.gov.uk/media/18032/SUS-
 Guidance-on-Data-Duplication-v19/pdf/SUS_Guidance_on_Data_Duplication_v1.9.pdf](http://www.hscic.gov.uk/media/18032/SUS-Guidance-on-Data-Duplication-v19/pdf/SUS_Guidance_on_Data_Duplication_v1.9.pdf)
- 36 The Health and Social Care Information Centre. Hospital Episode Statistics: Explanatory notes:
 maternity data, 2007-08. 2010. [http://www.hscic.gov.uk/catalogue/PUB01698/nhs-mate-
 eng-2007-2008-rep.pdf](http://www.hscic.gov.uk/catalogue/PUB01698/nhs-mate-eng-2007-2008-rep.pdf)
- 37 Abrahams C. Linking HES maternity records with ONS birth records. *Heal Stat Q* 2002;;22–30.
- 38 Office for National Statistics. Child mortality statistics: Childhood, infant and perinatal.
 2015.[http://www.ons.gov.uk/ons/rel/vsob1/child-mortality-statistics--childhood--infant-and-
 perinatal/index.html](http://www.ons.gov.uk/ons/rel/vsob1/child-mortality-statistics--childhood--infant-and-perinatal/index.html)

Table 1: Availability of data by birth dataset

Fields of information	HES deliveries	HES births	ONS births	NN4B
Birth status	✓	✓	✓	✓
Birth weight	✓	✓	✓	✓
Sex of baby	✓	✓	✓	✓
Mothers date of birth	✓	✓	*	*
Residential postcode	✓	✓	✓	✓
Residential address			✓	✓
Gestational age (still)	✓	✓	✓	✓
Gestational age (live)	✓	✓		✓
Ethnicity of mother	✓			
Ethnicity of baby		✓		✓
Previous pregnancies	✓	✓		
Previous children			✓	✓
Delivery method	✓	✓		
Delivery place	✓	✓		

*Although mothers date of birth is recorded in the ONS births dataset to comply with legislation on disclosure SAHSU researchers are only provided with two year maternal age bands.

Table 2: Missing data in selected birth outcome variables (2002-2010)

Variable	HES deliveries		HES births		ONS births		NN4B births ¹	
	n missing	% missing	n missing	% missing	n missing	% missing	n missing	% missing
Total births or deliveries (2002-2010)	5,545,905		5,534,194		5,727,407		3,333,154	
Maternal age	14605	0.3	212862 0	38.5	0	0	2	0
Plurality	905313	16.3	156391 8	28.3	N/A ²		N/A ²	
Residential postcode	36185	0.7	408807 1	73.9	318	0	885	0.0
Region (GOR)	32245	0.6	408657 4	73.8	0	0	6300	0.2
Carstairs score ³	49526	1	347363 7	71.4	8625	0.2	6300	0.2
Birth status (live/still)	154041 2	27.8	207239 3	37.4	N/A ⁴		N/A ⁴	
Birth weight ⁵	130020 2	23.4	208554 9	37.7	36224	0.6	20327	0.6
Sex of baby	132637 3	23.9	540904 8	97.7	0	0	2	0
Parity (previous children) ⁶	NI		NI		248018 0	48.4	NI	
Mothers' ethnicity	943329	17.0	NI		NI		NI	
Babies' ethnicity	NI		144115 4	26.0	NI		260112	7.8
Previous pregnancies	206153 9	37.2	539037 7	97.4	NI		NI	
Delivery method	129475 0	23.3	174296 0	31.5	NI		NI	
Delivery place	188190 1	33.9	221356 2	40.0	NI		NI	
Gestational age	220119 5	39.7	222761 8	40.3	N/A		32039	1

NI: Not included, NA: Not applicable

¹ 2006 to 2010 only

² ONS and NN4B codes multiple births as 1, all other births are considered singleton. Coverage is therefore 100%

³ No data for 2010, results displayed for years 2002 to 2009, NN4B 2006-2010

⁴ ONS tables are split into stillbirths and live births. Coverage is therefore 100%

⁵ Valid birth weights for HES include 1-7000g, for ONS 1-9997g and for NN4B 200-5000g

⁶ Applies to live and stillbirths within marriage only

Table 3: Lower layer Super Output Area level comparison of live births, stillbirths and average birth weight by selected government office regions 2002-2010

	Live births		Stillbirths		Continuous birth weight*	
	HES deliveries	ONS births	HES deliveries	ONS births	HES deliveries	ONS births
National (LSOA's=32,470)						
Mean	122.8	175.1	0.9	0.9	3368.6	3375.0
Standard deviation	75.6	80.8	2.6	1.2	121.9	94.0
Median	109	160	0	1	3378.4	3383.9
IQR	70-159	119-214	0-1	0-1	3306.8-3443.5	3316.4-3441.8
5-95 th percentiles	30-262	79-321	0-3	0-3	3185.0-3529.6	3208.4-3514.8
Min-Max	0-1491	3-1940	0-63	0-12	606-4600	2961.6-3734.1
North East (LSOA's = 1656)						
Mean	137.8	156.2	0.6	0.8	3380.1	3384.9
Standard deviation	59.4	67.2	0.9	1.0	86.8	84.7
Median	128	145	0	1	3382.1	3386.4
IQR	97-168	111-188	0-1	0-1	3321.6-3442.2	3329.0-3443.8
5-95 th percentiles	64-240	72-279	0-2	0-3	3231.6-3521.5	3241.6-3521.8
Min-Max	21-792	26-952	0-5	0-7	3020.0-3651.4	3072.8-3663.4
London (LSOA's=4765)						
Mean	156.2	226.7	0.9	1.4	3324.0	3321.6
Standard deviation	80.8	77.7	1.2	1.4	92.5	85.3
Median	149	220	1	1	3322.7	3321.4
IQR	94-207	174-269	0-1	0-2	3265.0-3385.3	3264.9-3380.9
5-95 th percentiles	41-298	114-359	0-3	0-4	3170.0-3472.1	3178.4-3461.9
Min-Max	2-674	3-1082	0-9	0-10	2971.3-3708.0	3013.1-3580.7

* Live singletons