National Maternity and Perinatal Audit

Technical Report

Linking the National Maternity and Perinatal Audit Data Set
to the National Neonatal Research Database for 2015/16

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE
RCPCH Royal College of Paediatrics and Child Health Leading the way in Children's Health
THE ROYAL COLLEGE OF MIDWIVES
Royal College of Obstetricians & Gynaecologists
HQIP Healthcare Quality Improvement Partnership

National Maternity and Perinatal Audit

Technical Report

Linking the National Maternity and Perinatal Audit Data Set to the National Neonatal Research Database for 2015/16

The National Maternity and Perinatal Audit (NMPA) is led by the Royal College of Obstetricians and Gynaecologists (RCOG) in partnership with the Royal College of Midwives (RCM), the Royal College of Paediatrics and Child Health (RCPCH) and the London School of Hygiene & Tropical Medicine (LSHTM)

This report was prepared by the NMPA project team

Lead author: Dr Harriet Aughey, NMPA Clinical Fellow (Neonatology)
Ms Andrea Blotkamp, NMPA Clinical Fellow (Midwifery)
Dr Fran Carroll, NMPA Research Fellow
Dr David Cromwell, NMPA Senior Methodological Advisor
Dr Ipek Gurol-Urganci, NMPA Senior Methodological Advisor
Dr Tina Harris, NMPA Senior Clinical Lead (Midwifery)
Dr Jane Hawdon, NMPA Senior Clinical Lead (Neonatology)
Dr Jen Jardine, NMPA Clinical Fellow (Obstetrics)
Dr Hannah Knight, NMPA Audit Lead
Dr Lindsey Mamza, NMPA Data Manager
Ms Natalie Moitt, NMPA Statistician
Dr Dharmintra Pasupathy, NMPA Senior Clinical Lead (Obstetrics)
Prof. Jan van der Meulen, NMPA Senior Methodologist (Chair)

NMPA neonatal advisory group members

Dr Breidge Boyle, Neonatal Nurses Association
Dr Chris Gale, Independent expert
Heidi Green, Wales representative
Dr Lesley Jackson, Scotland representative
Michele Upton, NHS Improvement
Rachel Winch, NNAP project manager
Dr Sam Oddie, NNAP Clinical Lead
Dr Steve Wardle, BAPM representative
Susanne Sweeney, Neonatal Operational Delivery Network
Zoe Chivers and Debbie Bezalel, Bliss

Please cite as: Aughey H, NMPA Project Team. *Technical Report: Linking the National Maternity and Perinatal Audit Data Set to the National Neonatal Research Database for 2015/16*. London: RCOG; 2019.

Contents

Tables and figures

Tables

Figures

Acknowledgements

We are very grateful to the midwives, nurses, doctors, maternity support workers, ward clerks, coders and data analysts who enter data into electronic maternity, neonatal care and patient systems on a daily basis.

The additional data for this project have come to us from the National Neonatal Research Database (NNRD). We would like to thank the team at the NNRD for their support in this process.

We are grateful to the advisory group for this project and their clinical insights, which will continue to drive this work.

Executive summary

Introduction

The National Maternity and Perinatal Audit (NMPA) is a national audit of the NHS maternity services across England, Scotland and Wales, commissioned in July 2016 by the Healthcare Quality Improvement Partnership (HQIP) on behalf of NHS England, the Welsh Government and the Health Department of the Scottish Government. The NMPA is led by the Royal College of Obstetricians and Gynaecologists (RCOG) in partnership with the Royal College of Midwives (RCM), the Royal College of Paediatrics and Child Health (RCPCH) and the London School of Hygiene & Tropical Medicine (LSHTM).

The overarching aim of the NMPA is to produce high-quality information about NHS maternity and neonatal services which can be used by providers, commissioners and users of the services to benchmark against national standards and recommendations where these exist, and to identify good practice and areas for improvement in the care of women and babies.

This short report from the NMPA explores the feasibility of linking the NMPA data set, which contains data relating to the majority of women who give birth, to the National Neonatal Research Database (NNRD), which contains detailed information about the majority of babies admitted to a neonatal unit. This feasibility study is limited to linkage between neonatal records and maternity records in England, as a pilot for developing this linkage across England, Scotland and Wales.

The linkage of neonatal data to maternity data offers many potential advantages. In particular, it allows the exploration of associations between maternal antenatal and intrapartum factors and neonatal outcomes. It also offers the potential to use one or more neonatal outcomes, such as admission to neonatal care, as an outcome measure of maternity care, and to explore variation in neonatal outcomes between maternity settings.

The purpose of this report is to describe the feasibility of linking the NMPA data set with the NNRD data set. It describes the technical process of linking these data sets and explores whether this linked data set can be used on an annual basis to construct clinically relevant measures of maternity care.

Methods

The NMPA holds data for 611 959 babies born in English NHS maternity services between 1 April 2015 and 31 March 2016. The NNRD contains clinical information for all babies in England admitted to neonatal care in the NHS. Records are derived from information entered into the clinical record system BadgerNet and are used by the Neonatal Data Analysis Unit (NDAU) to construct the NNRD.

An NNRD extract was supplied by NDAU to the NMPA. After cleaning and reshaping, this data extract contained 84 275 babies. The NMPA team then linked the two data sets using a deterministic linkage process and undertook validation checks by replicating measures produced by other national audit projects.

Provisional analyses of the linked data set were performed in order to evaluate the feasibility of producing clinical outcome measures using the linked data set.

Key findings

It is possible to link the NNRD and NMPA data sets using available identifying information, with an effective linkage rate of 96.7%. This represents an excellent level of linkage.

Validation checks identified that, in the process of generating the extract, NDAU had not included the records of approximately 8250 babies in the NNRD extract provided to the NMPA. This prevented further analysis, which will only be undertaken once the linkage process has been repeated using a complete NNRD extract.

Although potential clinical measures were identified, the NMPA was unable to generate these using the current linked data owing to the incomplete NNRD extract received.

The NMPA intends to re-link a complete NNRD data extract to the NMPA database and produce clinical outcome measures for publication in 2019.

The feasibility of future linkage will be determined by the ascertainment of the Maternity Services Data Set (MSDS). The utility of undertaking linkage between the NMPA and MSDS will be reviewed on an annual basis as MSDS ascertainment improves.

Recommendations

Recommendations for organisations requesting neonatal data extracts

1 Requests for extracts should be as specific and detailed as possible to ensure clarity about what is required from organisations providing data extracts.

2 Organisations that request data extracts from third parties such as NDAU should request written documentation or code, including details of any assumptions made, to explain how the extract was generated. This will ensure transparency and reproducibility of the data extract.

3 Where there is a desire for alignment between projects making use of the same neonatal data set, and particularly where similar clinical outcome measures are used, this should be clearly specified in the data request and discussed before data extraction. This will reduce the risk of publication of inconsistent results between the projects.

Recommendations for organisations supplying neonatal data extracts

4 Organisations supplying neonatal data extracts should provide complete data extracts, as requested, to allow for accurate analysis of these extracts by requesting organisations.

5 Organisations should provide details of the extraction process and an explanation of any assumptions made or validation checks performed. Any instances of dropped records due to a priori assumptions should be fully explained in writing and ideally discussed with the recipients beforehand.

Abbreviations and glossary

ATAIN	Avoiding Term Admissions into Neonatal Units (ATAIN) is a programme of work sponsored by NHS Improvement to reduce harm leading to avoidable admission to a neonatal unit for infants born at term, i.e. at ≥ 37^{+0} weeks of gestation. The programme is led by clinical experts from a range of organisations, including NHS Improvement, who work together to offer system-wide insights, practice points and evidence for healthcare organisations and professionals.
BadgerNet	A proprietary electronic medical record system containing neonatal care records, used by hospitals providing neonatal care in England, Scotland and Wales.
Case ascertainment	The proportion of cases captured in a data set when compared with the 'true' total or gold standard (for example, Office for National Statistics (ONS)).
CleverMed	A medical software company that runs the BadgerNet neonatal electronic record system.
Data cleaning	The process of detecting and correcting or removing records that are inaccurate, for example records that are incorrect, incomplete, improperly formatted or duplicated.
Deterministic linkage	Matching records between databases by comparing one or more fields that uniquely identify an individual (e.g. an individual's study ID) between the records; a link is made if they all agree.
HES	Hospital Episode Statistics (administered by NHS Digital), a data set containing details of admissions to all NHS hospitals in England (as well as outpatient appointments and accident & emergency attendances).
Linkage	The joining of two or more separate data sets to enrich the information on the individuals in the data sets. Data sets can be joined using (pseudo) identifiers and/or a combination of other fields to match the records from the separate data sets.
LSHTM	The London School of Hygiene & Tropical Medicine.
MIS	Maternity information system, an IT system containing maternity records with details on bookings and births.
MSDS	Maternity Services Data Set, administered by NHS Digital, which contains extracts from English NHS hospitals' MISs.
NDAU	Neonatal Data Analysis Unit, based at Imperial College London. This organisation administers the NNRD.
NNAP	National Neonatal Audit Programme, a national audit of care received by babies admitted to neonatal units in England, Wales and Scotland, whose analysis is performed using NNRD data. The NNAP is commissioned by the Healthcare Quality Improvement Partnership (HQIP), funded by NHS England, the Scottish Government and the Welsh Government, and delivered by the Royal College of Paediatrics and Child Health (RCPCH).
NNRD	National Neonatal Research Database, administered by the Neonatal Data Analysis Unit (NDAU), which uses data collected via the BadgerNet neonatal electronic record system. This database contains details of admissions to neonatal care in participating NHS hospitals in England, Scotland and Wales.
NNU	Neonatal unit, a hospital ward for the provision of specialist care to neonates.
NTC	Neonatal transitional care (NTC) supports resident mothers as primary care providers for their babies with care requirements in excess of normal newborn care, but who do not require admission to a neonatal unit (NNU).
ONS	Office for National Statistics
Probabilistic linkage	Matching records between databases by comparing multiple fields that may not uniquely identify an individual (e.g. an individual's first and last name).
RCOG	Royal College of Obstetricians and Gynaecologists
RCM	Royal College of Midwives
RCPCH	Royal College of Paediatrics and Child Health

Introduction

The National Maternity and Perinatal Audit

The National Maternity and Perinatal Audit (NMPA) is a national audit of NHS maternity services across England, Scotland and Wales, commissioned in July 2016 by the Healthcare Quality Improvement Partnership (HQIP) on behalf of NHS England, the Welsh Government and the Health Department of the Scottish Government. The NMPA is led by the Royal College of Obstetricians and Gynaecologists (RCOG) in partnership with the Royal College of Midwives (RCM), the Royal College of Paediatrics and Child Health (RCPCH) and the London School of Hygiene & Tropical Medicine (LSHTM).

The overarching aim of the NMPA is to produce high-quality information about NHS maternity and neonatal services that can be used by providers, commissioners and users of the services to benchmark against national standards and recommendations where these exist, and to identify good practice and areas for improvement in the care of women and babies. The NMPA consists of three separate but related elements:

- nn organisational survey of maternity and neonatal care in England, Scotland and Wales providing an overview of care provision and of services and options available to women
- a continuous clinical audit of a number of key measures to identify unexpected variation between service providers or regions
- a programme of periodic 'sprint' audits to evaluate the feasibility of linkage of the NMPA data set with other information sources for specific topics. The linkage of neonatal and maternity data discussed in this report is one of these feasibility studies.

Introduction to this feasibility study

This technical report is a summary of the work undertaken by the NMPA team to evaluate the feasibility of linking the NMPA data set, which contains data relating to the majority of women who give birth, to the National Neonatal Research Database (NNRD), which contains detailed information about the majority of babies admitted to a neonatal unit.

The data sources and methodology including the linkage process are described, as well as the strengths and potential uses of this linked data set. This report also describes the challenges encountered during the linkage and analysis processes, and the resulting limitations relating to the development of clinical measures. Finally, this report offers recommendations for organisations requesting data extracts and for organisations supplying data extracts.

This report relates to data for women giving birth and babies born in England in the financial year 1 April 2015 to 31 March 2016. The wider NMPA reports on England, Scotland and Wales over the same time period. However, systems and processes differ considerably between these countries, so it was decided to limit this data linkage feasibility study to one country in the first instance.

Rationale for this feasibility study

Linking maternal and neonatal data offers a unique opportunity to examine the maternal antecedents of neonatal outcomes and to determine the impact of aspects of maternal care on babies.

The NMPA data set[1] contains information about antenatal, intrapartum and postnatal care of the majority of mothers giving birth in England, Scotland and Wales. It also contains limited information about their babies, such as gestational age, birthweight and Apgar scores. Therefore, analysis of the NMPA data set alone can only provide information relating to a limited number of neonatal outcomes.*

The NNRD contains extensive information about the majority of babies admitted to neonatal care in England, Scotland and Wales,† including demographic data, diagnoses and data about specialist neonatal care processes. Therefore, analysis of the NNRD alone can give detailed neonatal information, for example the proportion of babies who are reported to have encephalopathy,[3] out of a cohort limited to babies admitted to neonatal care.

Linking these two data sets offers several potential advantages:

1 It provides information about all babies born in England. This allows analysts to accurately calculate proportions of the neonatal outcomes available in the NNRD out of all babies born (for example, rates of admissions of term babies to a neonatal unit).‡ Without linkage, calculation of such rates can only be done by comparing one data set with another. As data sets invariably have missing records, but not the same missing records, this results in inaccurate rates overall. Linkage ensures that we are comparing like with like and results in more accurate rates overall.

2 It allows exploration of associations between maternal, antenatal and intrapartum factors and neonatal outcomes.

3 It offers the potential for using one or more neonatal outcomes as outcome measures of maternity care and to explore variation in neonatal outcomes between maternity settings.

Objectives

The objectives of this project are:

1 to evaluate the feasibility of linking the NNRD to the NMPA data set

2 to explore whether this linked data set can be used to construct clinically relevant outcome and process measures

3 to evaluate the feasibility of repeating this linkage on an annual basis with a view to including the resulting neonatal clinical measures in the annually published NMPA continuous clinical audit.

* This report relates to data for women and babies born in England, but the wider NMPA reports on England, Scotland and Wales.

† Since 2012, 100% of English and Welsh neonatal units contribute to NNRD; since 2018, 100% of Scottish units contribute to NNRD.[2]

‡ The case ascertainment of the NMPA dataset for 1 April 2015 to 31 March 2016 is 92% of all births as captured by the Office for National Statistics (ONS).

Clinical outcome measures

Options for potential clinical outcome measures were agreed following consultation with the NMPA neonatal advisory group. This advisory group consisted of 22 members, representing a range of professional and third sector organisations as well as members of the existing NMPA team (the members are listed on page ii). The advisory group was convened to provide expert clinical input and interpretive support for proposed clinical measures. The following clinical outcome measures were considered to be feasible within the constraints of the data items available and also clinically relevant for both neonatal and maternity care.

Four proposed outcome measures were developed:

1 What proportion of term babies are admitted to a neonatal unit?

2 What proportion of late preterm babies are admitted to a neonatal unit?

3 What proportion of babies are reported to have an encephalopathy?

4 What proportion of term babies receive mechanical ventilation?

The first two measures would take advantage of the linked data set to show variation in admission rates for term and late preterm babies between trusts, which in turn could be used by individual trusts to benchmark and compare with other trusts.

The final two measures examine the proportion of babies that are recorded as having encephalopathy* and the proportion of term babies receiving mechanical ventilation. Each of these measures takes a very specific adverse neonatal outcome recorded in the NNRD data set and allows these outcomes to be presented as a proportion of all babies. These two adverse neonatal outcomes could potentially be considered proxy measures for the quality of perinatal care.

The Avoiding Term Admissions into Neonatal Units (ATAIN) programme, as part of work to reduce unnecessary admissions of term babies to neonatal units, has published admission rates of term babies to neonatal units in previous years.[4] The National Neonatal Audit Programme (NNAP) also reports a related measure that has some overlap with the second proposed NMPA measure for admission rates of late preterm babies.† The NMPA team is therefore aware of the importance of ensuring that any results produced through the analysis of a linked NMPA–NNRD data set are directly comparable with results produced by other national projects.

A more detailed description of the NMPA neonatal clinical measures agreed can be found in Appendix 1.

In addition to these measures, the NMPA neonatal advisory group was interested in the potential to use a linked data set to explore associations between maternal factors that are beyond the influence of maternity services, such as obesity, and adverse neonatal outcomes, such as admission to a neonatal unit. While the association between such maternal factors and neonatal outcomes cannot be considered to reflect the quality of maternity care, describing these associations may nonetheless result in profound public health and policy messages. Therefore, it was agreed that these associations would be reported at national level.

* Encephalopathy means any acute disorder of the brain. In the context of term neonates, this is most commonly hypoxic ischaemic encephalopathy, which is an acute disorder that occurs owing to a lack of oxygen to the fetal or neonatal brain.

† NNAP measure 2018 – Minimising inappropriate separation of mother and term baby: 'For babies with a birth gestation of less than or equal to 37 weeks, who did not have any surgery or a transfer during any admission: How many Special Care or Normal Care days were provided when oxygen was not administered?'[3]

Methods

Data sources

Information about births: the National Maternity and Perinatal Audit (NMPA) data set

The vast majority of NHS trusts in England with a maternity service use an electronic maternity information system (MIS) to capture detailed demographic and clinical information related to each woman and baby under their care. These data, entered by midwives and support staff in the antenatal clinic and/or maternity unit, typically cover antenatal booking through to birth and immediate postnatal care.

The NMPA team requested an extract from each trust's MIS and then cleaned and linked the data at record level to maternal Hospital Episode Statistics (HES) inpatient care records to allow the use of HES variables not present in MIS records. The resulting NMPA data set contains records of 611 959 babies born in England between 1 April 2015 and 31 March 2016, which represents 92% of total births registered by the Office for National Statistics (ONS) for the reporting period.[1]

Further detail about the NMPA data collection and analysis processes can be found in the NMPA's *Clinical Report 2017*.[1]

Information about neonatal admissions: the National Neonatal Research Database (NNRD)

Neonatal clinical data are entered by clinical staff into BadgerNet, an electronic medical records system used by the majority of NHS trusts providing neonatal care and administered by CleverMed. CleverMed supplies the data to the Neonatal Data Analysis Unit (NDAU), an independent academic unit based at Imperial College London. NDAU uses these neonatal data to construct the NNRD.[2]

The NNRD therefore contains clinical information captured in the course of neonatal care. NNRD data are made available for multiple purposes including health services evaluation and research. The NMPA team requested an extract from the NNRD including all babies born between 1 April 2015 and 31 March 2016.

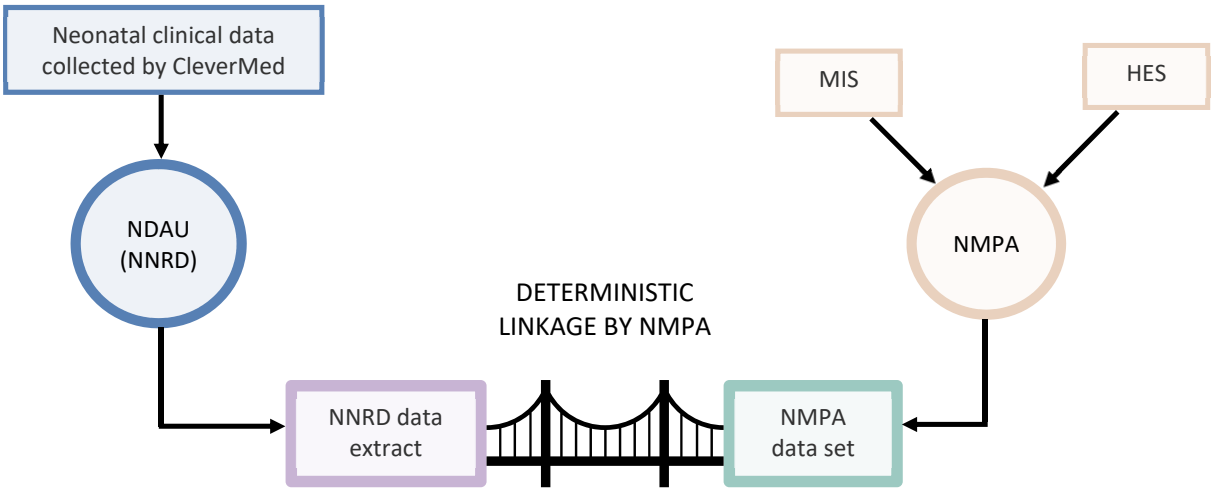

Figure 1 Data sources for the NMPA–NNRD linkage feasibility project

Cleaning the NNRD extract

The NMPA team received a data extract from the NNRD containing over 1 million records relating to the 1 April 2015 to 31 March 2016 financial year. The extract contained multiple records per day of care for each baby; the number of records for each baby ranged from 1 to 369 records, with a median of 4 records per baby. Before the NNRD extract could be linked to the NMPA data set, it was necessary to identify records belonging to individual babies and reshape the NNRD extract into a one-row-per-baby format.

Individual babies may be identified in the NNRD data set using pseudonymised identifiers (pseudo-IDs), of which there are several. The NMPA team has the advantage of having access to both the NNRD pseudo-IDs and personal identifier fields for babies (e.g. NHS number, date of birth or postcode). At present, NDAU does not have access to personal identifier fields. This meant that the NMPA team were in a position to perform checks, not available to the NDAU team, that allowed verification of the reliability of the NNRD pseudo-IDs. The results of this validation can be found in Appendix 2.

After the cleaning and reshaping process, the NMPA's NNRD extract contained 84 275 babies in the one-row-per-baby format.

Linkage methods

The linkage methods employed by the NMPA team* relied on comparing the variables listed in Table 1 between the data sets. The variables of interest were found to be well completed in both data sets.

Table 1 List of identifying variables present in both data sets

Variable	Completeness in NNRD	Completeness in NMPA
Baby's NHS number	98.2%	98.2%
Baby's date of birth	99.7%	99.6%
Baby's postcode	94.8%	90.2%
Baby's gender	99.9%	98.8%
Baby's birthweight	100.0%	97.2%
Number of infants (multiple births)	99.9%	98.5%
Birth order	99.9%	95.0%
Gestation at birth	99.9%	96.3%
Mother's NHS number	94.2%	99.6%
Mother's date of birth	93.8%	97.0%
Site of birth	98.4%	98.5%

These variables were separated from all other fields in both the cleaned NMPA and cleaned NNRD data sets. The NMPA and NNRD records were then linked using deterministic methods in a series of steps using different combinations of these identifying variables (Table 1). This process is depicted in Figure 2.

When linking NNRD records to NMPA records, it was found that a single NNRD record may link to multiple NMPA records. The converse was also true, in that a single NMPA record may link to multiple NNRD records. This can be explained as being due to duplicates in the data set, incorrect

* This is an abbreviated explanation of a primarily deterministic linkage process that was adapted specifically to these data. More information is available from the NMPA team on request.

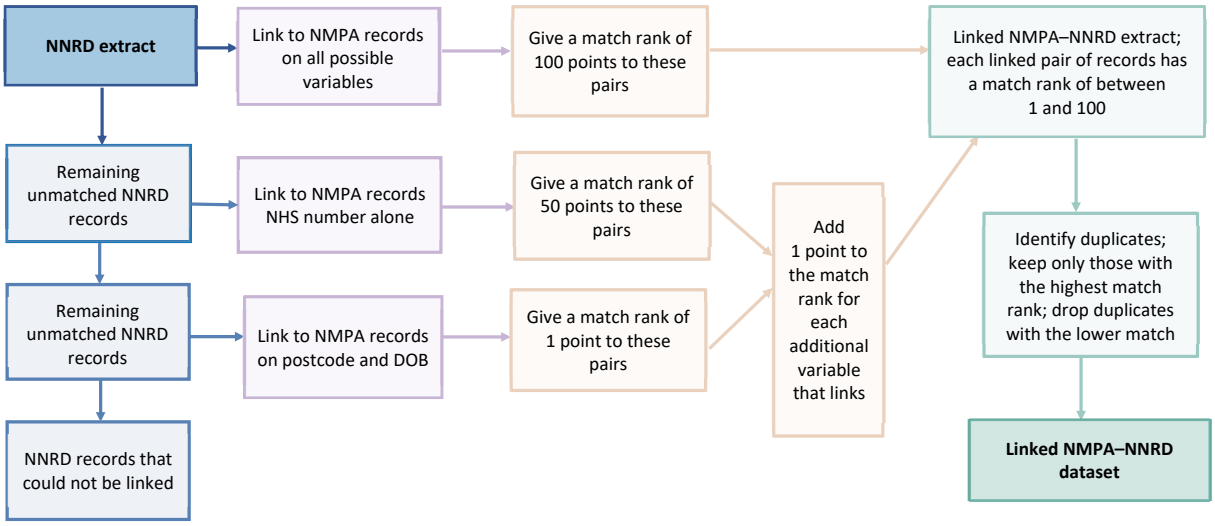

Figure 2 Linkage steps and assignment of match ranks

matches or two different babies genuinely sharing the same demographic data (i.e. postcode and date of birth).* Therefore, it was necessary to determine which pair of NMPA and NNRD records had the best-quality linkage, and was therefore most likely to truly belong to the same baby.

To do this, the NMPA applied a 'match rank'† to each pair of linked NMPA–NNRD records using the methodology shown in Figure 2. Where more than one NMPA record linked with an NNRD record, the paired NMPA–NNRD record with the highest match rank was considered to be the true match, and the pairs with the lower match rank were dropped.

The highest match rank was given to pairs of records that matched on all possible variables. Pairs that matched on NHS number alone were given the second highest match rank, and within this group, additional points were added to the match rank for each additional variable that also linked. Finally, pairs that matched on postcode and date of birth only were given the lowest match rank, and again, within this group, additional points were added to the match rank for each additional variable that also linked (Table 2).

Evaluation of the linkage

In order to fully evaluate the ascertainment of a linked NMPA–NNRD data set it would be necessary to compare it with a 'gold standard' linked data set, in which each baby receiving neonatal care is reliably identified. At present the NMPA team does not have access to such a data set and therefore this evaluation is not possible.

The NNRD can be considered the best available standard data set, since all babies born in England who receive neonatal unit care should be captured in the NNRD. Almost all of these should link to a maternity record, since all babies, including those admitted to neonatal care, will have been born to a mother receiving maternity care who should therefore have a maternity record. The exceptions are babies born outside of the English NHS but admitted to NHS neonatal care, or those babies born to mothers without a record in the NMPA.‡

* The most obvious example of this is multiple births.

† The match rank is a score given to a linked pair of records to denote the quality of the linkage. A higher match rank represents a better linkage.

‡ The NMPA data set contains only 92% of the total births in England for 1 April 2015 to 31 March 2016. Therefore, even with a perfect linkage methodology, some NNRD records would remain unlinked. The majority of NMPA records would not be expected to match with an NNRD record since the majority of babies born are not admitted to neonatal care.

Therefore, to quantify the quality of the NMPA linkage process, the following steps were undertaken:

1 An effective linkage rate was calculated that accounted for the incomplete NMPA ascertainment (92%). The effective linkage rate was calculated as the proportion of NNRD records that linked with an NMPA record multiplied by the case ascertainment of the NMPA data set.

2 The proportion of NNRD records that linked to an NMPA record was calculated and inspected by the trust of birth, and reasons for non-linkage were investigated.

3 The characteristics of linked and unlinked NNRD records were compared in order to investigate potential sources of bias.

Results

Linkage rates

Overall, 89% of the NNRD records were linked with an NMPA record. Given that the NMPA data set contains 92% of all births in England during this period, and that the NNRD data set contains close to 100% of all babies admitted to neonatal care, this equates to an effective linkage rate of 96.7%.*

Figure 3 shows the proportion of records in the NNRD extract for babies born in the 1 April 2015 to 31 March 2016 financial year that were matched and unmatched at the end of each linkage step.†

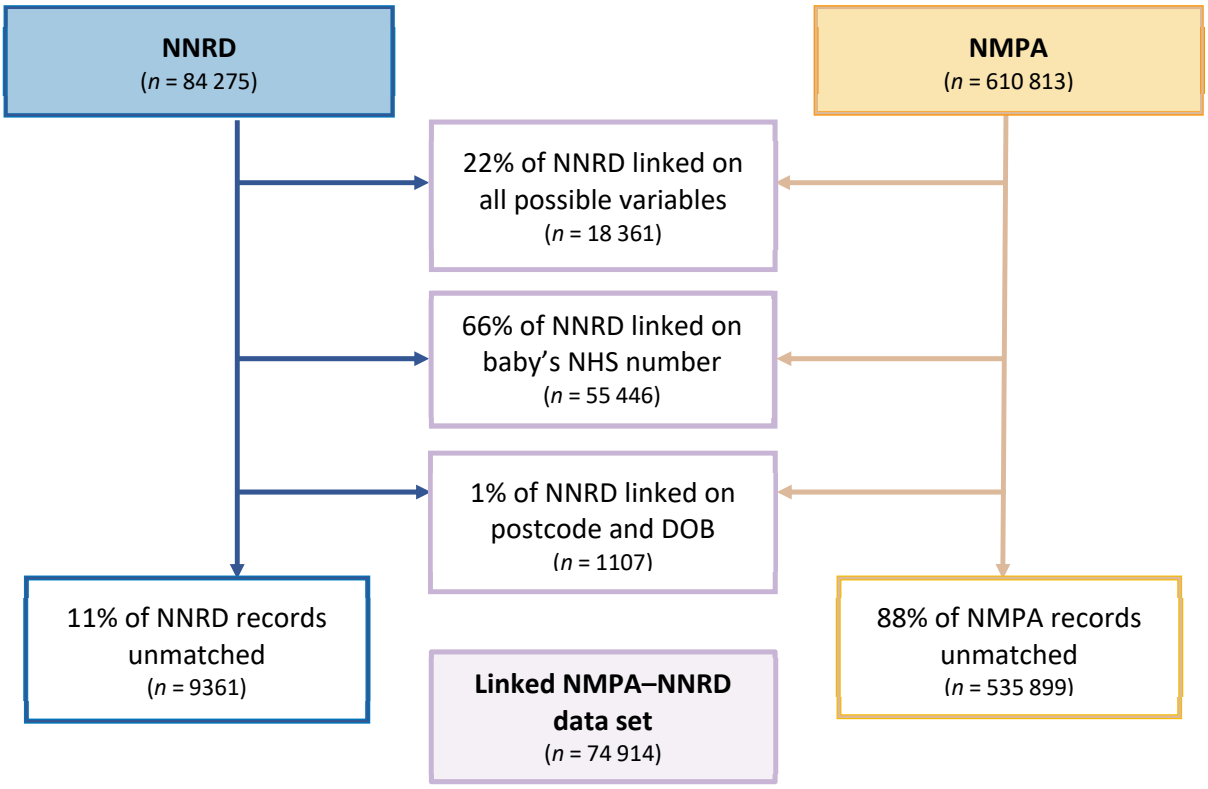

Figure 3 The number of NNRD records linked to an NMPA record at each step of the linkage process

Figure 4 illustrates the linkage rate in each individual NHS trust in England and shows the overall very high rates of linkage in the majority of trusts. Of the 134 trusts in England, 114 trusts (85%) had a linkage rate of more than 85%, 109 trusts (81%) had a linkage rate of more than 90% and 99 trusts (74%) had a linkage rate of more than 95%. Among NHS trusts with a linkage rate of more than 85%, only 1586 babies (2.3%) in the NNRD extract were not linked to NMPA records.

There are a small number of trusts (9 trusts, 6.7%) with a linkage rate of less than 1%: these are trusts that did not submit data extracts to the NMPA. Furthermore, in certain trusts, linkage rates are

* Calculated from 89/92 = 96.7%

† The majority of NMPA records did not match with an NNRD record – this was to be expected, as the majority of babies in the NMPA data set were not admitted to neonatal care.

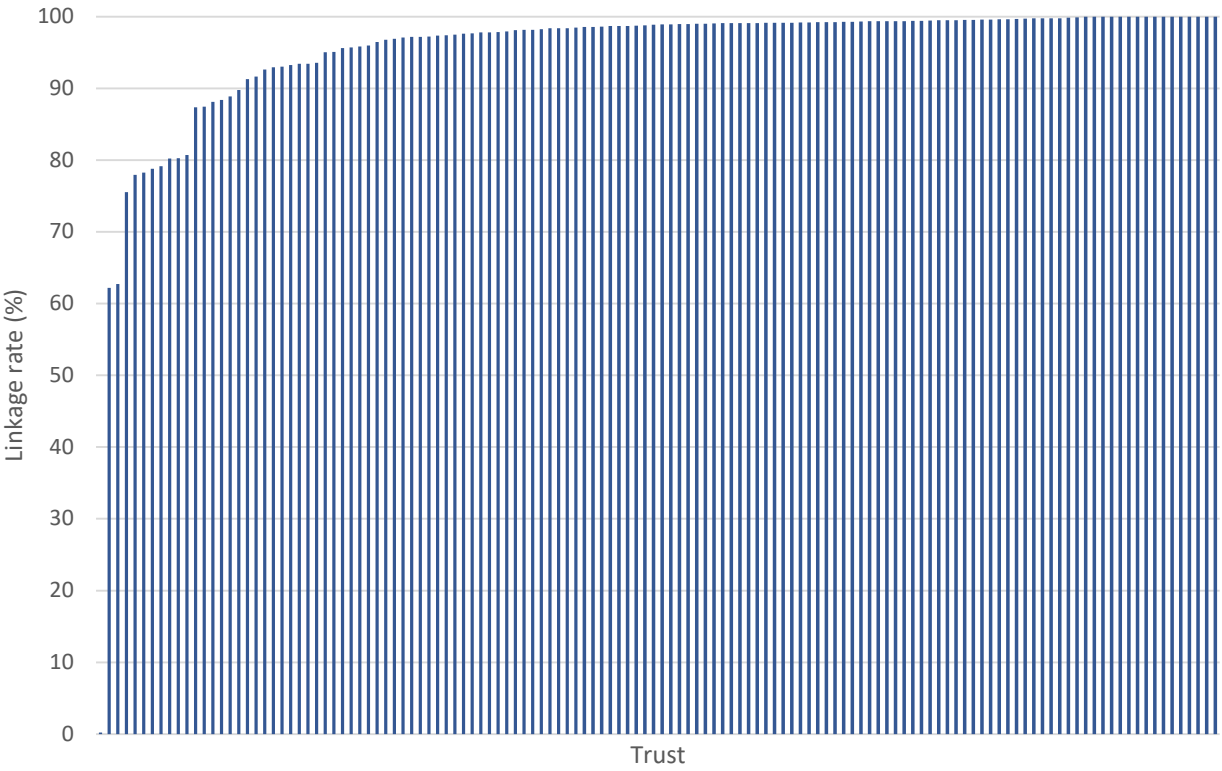

Figure 4 Linkage rate (% of NNRD records linked to an NMPA record) by individual trust in England

restricted by the fact that identifiers are incorrect, incomplete or missing within the NMPA and/or NNRD data sets.

Characteristics of linked and unlinked NNRD records

When comparing the characteristics of NNRD records that were linked with those which were not linked (Table 2), it was found that NNRD records that did not link belonged, on average, to babies with lower birthweight and shorter gestational age than the linked records. This difference is more pronounced when comparing linked and unlinked records within trusts with a higher overall linkage rate (e.g. limited to trusts with a linkage rate of more than 85%).

Table 2 Characteristics of linked versus unlinked NNRD records

All English babies in NNRD extract (all trusts)	Linked records	Unlinked records
Mean birthweight	2839 g	2698 g
Mean gestational age	260 days	255 days
Number of babies	74 892 babies	9202 babies
All English babies in NNRD extract (trusts with >85% linkage)	**Linked records**	**Unlinked records**
Mean birthweight	2828 g	2351 g
Mean gestational age	260 days	246 days
Number of babies	66 621 babies	1586 babies

Validation of the linked data set

This is the first time that the NMPA data set has been linked to an NNRD extract; however, the NNRD is used by many other projects to analyse neonatal data. Therefore, in order to validate the newly linked NMPA–NNRD data set, the NMPA team performed analyses on this data set with the intention of replicating some of the measures produced by these other projects (NNAP* and ATAIN†).

This yielded initial results that were more discrepant than expected. A certain degree of discrepancy is expected due to:

1 use of different denominator data − the NMPA uses the NMPA data set as a denominator, rather than the ONS birth register

2 exclusion of unlinked records – NNRD records that could not be linked to NMPA records are not included in this current analysis whereas they would be included in other projects using the same (unlinked) NNRD data.

Upon further investigation it became apparent that the reason for this was that approximately 110 000 records, equating to approximately 8250 babies, were missing from the NNRD extract that the NMPA received from NDAU. During the preparation of the data extract,‡ a data cleaning step was taken by the NDAU team that resulted in all records being dropped where the care episode date and time was earlier than the date and time of birth or admission to neonatal care. In some instances where a baby had only one row of daily data, this baby was effectively removed from the data set. This data cleaning step was not applied during the preparation of NNRD data supplied to other national projects and therefore these records were not dropped from the extracts received by ATAIN and NNAP during the same time period, meaning that the analyses conducted by these projects included approximately 8250 more babies than those of the NMPA.

The issue with the care episode date and time in the dropped records is a data quality issue limited to a single data field. It is therefore likely that the missing records represent true neonatal admissions and that the NMPA's initial NNRD data set was incomplete.

Feasibility of using a linked NMPA–NNRD data set to derive clinical measures

Owing to concerns relating to the completeness of the NNRD extract received, the NMPA team, in consultation with stakeholders, decided that full analysis of clinical measures is not appropriate until a complete data extract is received and linked.

In order to evaluate the feasibility of producing the clinical measures previously described, some provisional analyses of the initial linked data set were performed. The following results relating to one clinical measure only are therefore intended to illustrate the feasibility of producing clinically relevant measures and are not intended for clinical use.

The NMPA team analysed the linked NMPA–NNRD data set to determine the proportion of term babies (born between 37^{+0} and 42^{+6} weeks of gestation) admitted to a neonatal unit in England during the financial year 1 April 2015 to 31 March 2016.

* The NNAP measure reproduced was parental consultation within 24 hours of admission: 'Is there a documented consultation with parents by a senior member of the neonatal team within 24 hours of admission?'[3]

† The ATAIN measure reproduced was the proportion of term babies admitted to neonatal care in England in the 1 April 2015 to 31 March 2016 financial year (unpublished data).

‡ It was communicated by NDAU analysts that a data cleaning step had been applied to an initial data extract that contained data for three financial years. This data cleaning step removed 24 754 babies (329 855 records), which resulted in an estimated 110 000 missing records or 8250 missing babies for the 1 April 2015 to 31 March 2016 financial year (329 855/3 and 24 757/3 respectively).

After data quality checks were applied, this analysis found a term admission rate of 4.7%. This is almost certainly an underestimate of the true admission rate of term babies in England in the financial year 1 April 2015 to 31 March 2016. Approximately 110 000 records had been dropped from the NNRD extract. Collaboration with NDAU determined that this represents approximately 8250 babies, of which 8200* were term babies. Therefore, if these 8200 term babies had all been admitted to a neonatal unit and were able to be linked, this would result in a term admission rate of 6.6%.†

Further detail relating to clinical measures can be found in Appendix 1.

* This figure was estimated by NDAU analysts, using NDAU's processes for identifying individual babies. The number of term babies in the 2015/16 time period in the original extract (from which records had been dropped) was compared with the number of term babies that would be present in the 2015/16 time period if the data cleaning step had not been applied (no babies dropped).

† If some or all of these term babies had received only neonatal transitional care (NTC) or postnatal care, and had therefore not been admitted to a neonatal unit, this number would be lower.

Discussion

Success of the data linkage process

Objective	To evaluate the feasibility of linking the NNRD to the NMPA data set.
Results	This study achieved an effective linkage rate between the NMPA and NNRD of 96.7%. This represents an excellent level of linkage and therefore the NMPA team conclude that linkage between these two data sets is feasible.

This feasibility study found that it was possible to link the NMPA and NNRD with an estimated effective linkage rate of 96.7%.* This represents an excellent level of linkage† between these two data sets. Therefore, it is likely that any conclusions drawn from future analysis of an NMPA–NNRD data set linked in this way will be robust.

Analysis of the linkage rate by individual NHS trust showed that the majority of individual trusts had either very high or very low linkage rates (Figure 4). This indicates that the main factor limiting the linkage of records was not the linkage methodology employed. Rather, it was due to incompleteness of identifiers in the two data sets and to the fact that approximately 8% of births are not captured in the NMPA data set. Therefore, the ascertainment of a linked NMPA–NNRD data set could be improved further through:

1 increasing the case ascertainment of the NMPA data set

2 improving the completeness of identifiers in both the NMPA and NNRD data sets.

The fact that small, preterm babies are less likely to be represented in the linked NMPA–NNRD data set is a potential source of bias. This issue is also seen in other studies.[7] A possible clinical explanation for this is that small, preterm babies are more likely to be unwell and therefore may be rapidly admitted to a neonatal unit, which could lead to less complete entry of data relating to these babies into maternity data systems. It should be noted that using the linked data set to specifically study small or preterm babies should be approached with caution.

Clinical outcome measures

Objective	To explore whether this linked data set could be used to construct clinically relevant outcome and process measures.
Results	Although potential clinical measures were identified, the NMPA team were unable to generate these owing to the receipt of an incomplete NNRD extract.

Provisional analysis of an initial linked NMPA–NNRD data set demonstrates that it is likely to be feasible to use this data set to produce clinical measures. However, owing to the incompleteness of the NNRD extract (described on p. 10), the NMPA team did not perform any further analysis or additional validation of results. The intention is ultimately to produce descriptive statistics and clinical outcome measures based on a linked NMPA–NNRD data set that includes the 110 000 records (relating

* Effective linkage rate allowing for the NMPA 92% case ascertainment for the the 1 April 2015 to 31 March 2016 financial year.

† There is no agreed standard for a 'good' level of linkage; however, other studies have considered linkage rates of more than 80% as 'good' and 'high quality'.[5,6]

to 8250 babies) that had been dropped from the initial NNRD extract received by the NMPA team. The intended date of publication for these measures is 2019.

Linking the Maternity Services Data Set (MSDS) to the NNRD in future years

Objective	To evaluate the feasibility of repeating this linkage on an annual basis with a view to including resulting neonatal clinical measures in the annually published NMPA continuous clinical audit.
Results	Future linkage will involve linkage of the NNRD to the Maternity Services Data Set (MSDS). The MSDS ascertainment is currently too low to make linkage worthwhile; however, this will be reviewed on an annual basis as MSDS ascertainment improves.

In England, a new Maternity Services Data Set (MSDS), managed by NHS Digital, has been developed to provide a data source that can inform how the quality of maternity services can be improved in the English NHS. However, this data set was not sufficiently complete to be used as a data source for the NMPA for the 1 April 2015 to 31 March 2016 financial year.

The first report for the NMPA using the MSDS in place of MIS data will be published in 2019 and will cover babies born between 1 April 2017 and 31 March 2018. Combining the figures within NHS Digital's Maternity Services Monthly Statistics reports, 503 017 babies reported to have been born in this period had records within the MSDS, from which the NMPA team estimates that the case ascertainment of the MSDS is approximately 77%.* The case ascertainment varies widely between months, for example with a case ascertainment of 66% in April 2017 and 85% in March 2018. Even if all variables needed for linkage had perfect data quality in the MSDS, the ascertainment of a linked data set would be limited by the case ascertainment of the MSDS.

Owing to the relatively low case ascertainment and completeness of the MSDS, the NMPA does not intend to link MSDS with NNRD data for babies born between 1 April 2017 and 31 March 2018. The NMPA team will reconsider the linkage of MSDS and NNRD data once the data quality and completeness of the MSDS are deemed sufficient to make this linkage worthwhile. This will be reviewed on an annual basis. As indicated earlier, there are indications that the MSDS case ascertainment is improving over time, so it is likely that in the future robust analysis will be feasible using linked MSDS–NNRD data.

* Using the number of live births (646 794) and the stillbirth rate (4.1 per 1000 babies) in England in 2017 published by the ONS gives an estimated total number of births of 649 457 in England in 2017. This number is used as an estimate for the total number of births in England in the 2017/18 financial year. This is divided by 365.25 and multiplied by the number of days in the specified month to derive the equivalent monthly estimate.

Key findings

It is possible to link the NNRD and NMPA data sets using available identifying information, with an effective linkage rate of 96.7%. This represents an excellent level of linkage.

Validation checks identified that, in the process of generating the extract, NDAU had not included the records of approximately 8250 babies in the NNRD extract provided to the NMPA. This prevented further analysis, which will only be undertaken once the linkage process has been repeated using a complete NNRD extract.

Although potential clinical measures were identified, the NMPA was unable to generate these using the current linked data owing to the incomplete NNRD extract received.

The NMPA intends to re-link a complete NNRD data extract to the NMPA database and produce clinical outcome measures for publication in 2019.

The feasibility of future linkage will be determined by the ascertainment of the Maternity Services Data Set (MSDS). The utility of undertaking linkage between the NMPA and MSDS will be reviewed on an annual basis as MSDS ascertainment improves.

Recommendations for organisations requesting neonatal data extracts

1 Requests for extracts should be as specific and detailed as possible to ensure clarity about what is required from organisations providing data extracts.

2 Organisations that request data extracts from third parties such as NDAU should request written documentation or code, including details of any assumptions made, to explain how the extract was generated. This will ensure transparency and reproducibility of the data extract.

3 Where there is a desire for alignment between projects making use of the same neonatal data set, and particularly where similar clinical outcome measures are used, this should be clearly specified in the data request and discussed before data extraction. This will reduce the risk of publication of inconsistent results between the projects.

Recommendations for organisations supplying neonatal data extracts

4 Organisations supplying neonatal data extracts should provide complete data extracts, as requested, to allow for accurate analysis of these extracts by requesting organisations.

5 Organisations should provide details of the extraction process and an explanation of any assumptions made or validation checks performed. Any instances of dropped records due to a priori assumptions should be fully explained in writing and ideally discussed with the recipients beforehand.

References

1. NMPA Project Team. *National Maternity and Perinatal Audit: Clinical Report 2017 – Revised Version*. London: RCOG; 2018 [www.maternityaudit.org.uk/downloads/NMPA%20Clinical%20Report%202018.pdf].
2. Neonatal Data Analysis Unit (NDAU), Imperial College London. *Neonatal data* [www.imperial.ac.uk/neonatal-data-analysis-unit/neonatal-data/].
3. National Neonatal Audit Programme (NNAP), Royal College of Paediatrics and Child Health. *NNAP Online* [https://nnap.rcpch.ac.uk/].
4. Battersby C, Michaelides S, Upton M, Rennie JM on behalf of the Jaundice Working Group of the Atain (Avoiding Term Admissions Into Neonatal units) programme, led by the Patient Safety team in NHS Improvement. Term admissions to neonatal units in England: a role for transitional care? A retrospective cohort study *BMJ Open* 2017;7:e016050 [https://bmjopen.bmj.com/content/7/5/e016050].
5. Dattani N, Datta-Nemdharry P, Macfarlane A. Linking maternity data for England, 2005-06: methods and data quality. *Health Stat Q* 2011;49:53–79.
6. Hammill BG, Hernandez AF, Peterson ED, Fonarow GC, Schulman KA, Curtis LH. Linking inpatient clinical registry data to Medicare claims data using indirect identifiers. *Am Heart J* 2009;157:995–1000 [www.ncbi.nlm.nih.gov/pmc/articles/PMC2732025/].
7. Harron K, Gilbert R, Cromwell D, van der Meulen J. Linking data for mothers and babies in de-identified electronic health data. *PLoS One* 11: e0164667 [www.ncbi.nlm.nih.gov/pmc/articles/PMC5072610/].

Appendix 1

Technical specification – NMPA neonatal morbidity measures using the NNRD

Below is the provisional technical specification for the analysis of a linked NMPA–NNRD data set to derive clinical measures. Details within this document are subject to change as necessary when analysis of a final data set is undertaken.

The following measures are dependent on the NMPA–NNRD linkage and will be only reported for England.

Proportion of babies born at term admitted to a neonatal unit

Rationale: To describe variation in rates of term admissions and allow for benchmarking. NMPA has the additional advantage of case-mix adjustment of maternal factors to allow for meaningful comparisons.

Data sources: NMPA, NNRD

Site types: All

Numerator: Number of liveborn singleton babies born between 37^{+0} and 42^{+6} weeks of gestation admitted to a neonatal unit (NNU)

Denominator: All liveborn singleton babies born between 37^{+0} and 42^{+6} weeks of gestation

Case-mix factors: Default

Level/s of reporting: Trust/board, regional, national

Proportion of babies born in the late preterm period (34^{+0} to 36^{+6} weeks) admitted to a neonatal unit

Rationale: To describing variation in rates of late preterm admissions

Data sources: NMPA, NNRD

Numerator: Number of liveborn singleton babies born between 34^{+0} and 36^{+6} weeks of gestation admitted to a NNU

Denominator: All liveborn singleton babies born between 34^{+0} and 36^{+6} weeks of gestation

Case-mix factors: Default

Level/s of reporting: Trust/board, regional, national

Proportion of babies born between 35^{+0} and 42^{+6} weeks of gestation who have an encephalopathy

Rationale: To describe variation in rates of neonatal encephalopathy with the aim of reflecting variations in obstetric care

Data sources: NMPA, NNRD

Numerator: Number of liveborn singleton babies born between 35^{+0} weeks and 42^{+6} weeks of gestation* admitted to an NNU for at least 72 hours, and with two or more of the following neurological signs recorded in the same daily data summary within the first 72 hours of life:

- Tone: abnormal
- Consciousness: lethargic or comatose
- Convulsions: yes

Denominator: Number of liveborn singleton babies born between 35^{+0} and 42^{+6} weeks of gestation

Case-mix factors: Default

Level/s of reporting: Trust/board, regional, national

Proportion of babies born at term who require mechanical ventilation in the first 72 hours of life

Rationale: Requiring mechanical ventilation at full term is a marker of an unwell baby, and in the first 3 days the most likely non-surgical causes are those related to birth. Demonstrating variation in rates of mechanical ventilation in term babies could be used as an outcome measure reflecting maternity care.

Data sources: NMPA, NNRD

Numerator: Number of liveborn singleton babies born between 37^{+0} and 42^{+6} weeks of gestation who have at least one recorded episode of mechanical ventilation within the first 72 hours of life

Denominator: Number of liveborn singleton babies born between 37^{+0} and 42^{+6} weeks of gestation

Case-mix factors: Default

Level/s of reporting: Trust/board, regional, national

* This is the same gestational age range used by the NNAP encephalopathy measure: 'Does an admitted baby born at 35 weeks gestational age or above have an encephalopathy within the first three full calendar days after birth?'

Appendix 2

Validation of NNRD pseudo-IDs

The NMPA team generated a new pseudo-ID (NMPA-pseudo-ID) for individual babies in the NNRD extract, derived using a combination of personal identifier fields and other distinguishing fields such as birthweight and site of birth. Defining an individual's record in this way is more robust than using any single field. For example, records with the same date of birth, mother's NHS number, postcode, birthweight, site code of delivery and gestational age are classified as belonging to the same baby, meaning that a record can be assigned to the corresponding baby if the baby's NHS number field is empty.

Using the NMPA-pseudo-ID to define individual babies, the NMPA team were able to investigate whether the NNRD pseudo-IDs ('Badger ID' and 'nationalidbabyanon') could be reliably used to identify individual babies within the NNRD data set. This process was then repeated for individual mothers within the data set. This analysis was performed on the NNRD data extract spanning the 1 April 2015 to 31 March 2016 financial year, which contained 84 275 individual babies.

Out of the 84 275 babies in the NMPA's NNRD extract:

- 63 babies have more than one 'Badger ID'
- 209 babies have non-unique 'Badger IDs'
- 64 babies have more than one value of 'nationalidbabyanon'
- 193 babies have a non-unique 'nationalidbabyanon'
- 48 mothers have more than one value of 'nationalidmotheranon'
- 303 mothers have non-unique 'nationalidmotheranon'
- 2316 mothers have no 'nationalidmotheranon'.

This shows that, out of the 84 275 babies in the extract, the NMPA team found only very small numbers that had problems with NNRD pseudo-IDs.

Therefore, the NMPA team concluded that NNRD pseudo-IDs are robust. The use of these pseudo-IDs to identify individual babies will introduce only very minimal bias and therefore these pseudo-IDs can be reliably used in analysis of the NNRD data set.

National Maternity and Perinatal Audit
Lindsay Stewart Centre for Audit and Clinical Informatics
Royal College of Obstetricians and Gynaecologists
27 Sussex Place, Regent's Park, London NW1 4RG
Email: nmpa@rcog.org.uk
Website: www.maternityaudit.org.uk
Twitter: @nmpa_audit

NMPA
National Maternity & Perinatal Audit